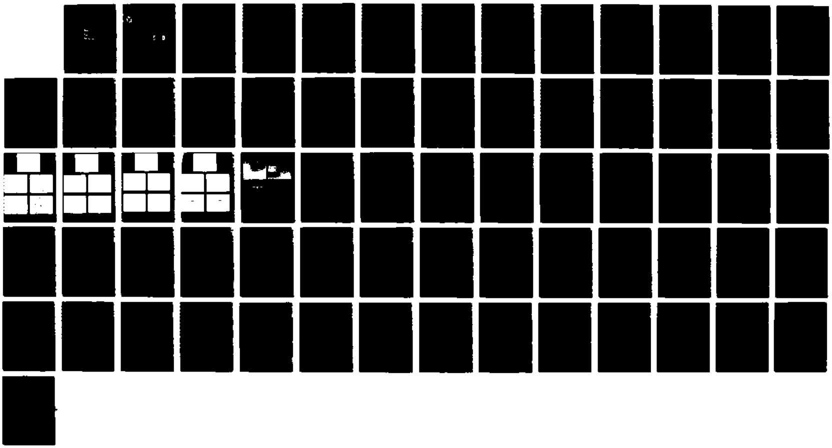
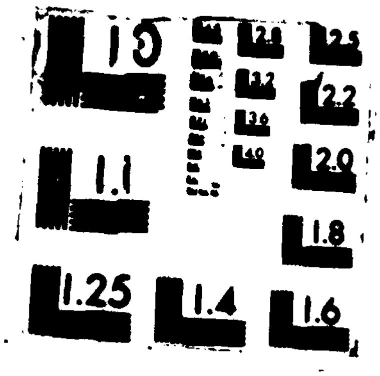
AD-A183 660 DEVELOPMENT OF AN OBJECTIVE STANDARDIZED TEST FOR VIDEO TELECONFERENCING(U) DELTA INFORMATION SYSTEMS INC HORSHAM PA FEB 87 NCS-TIB-87-7 DCA100-83-C-0047 1/1

UNCLASSIFIED F/G 25/3 NL




1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

(2)

AD-A183 660

NCS TIB 87-7

NATIONAL COMMUNICATIONS SYSTEM

TECHNICAL INFORMATION BULLETIN 87-7

DEVELOPMENT OF AN OBJECTIVE STANDARDIZED TEST FOR VIDEO TELECONFERENCING

FEBRUARY 1987

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NCS TIB 87-7	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Delta Information Systems 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Horsham Business Center, Bldg. 3 300 Welsh Road Horsham, PA 19044	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION National Communications System 8b. OFFICE SYMBOL (If applicable) NCS-TS	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DCA100-83-C-0047
8c. ADDRESS (City, State, and ZIP Code) Office of Technology & Standards Washington, D.C. 20305-2010	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: — PROJECT NO.: P00009 TASK NO.: 1 WORK UNIT ACCESSION NO.: —

11. TITLE (Include Security Classification)

Development of an Objective Standardized Test for Video Teleconferencing

12. PERSONAL AUTHOR(S)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM ____ TO ____	February 1987	67

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Video Teleconferencing Digital Video Codecs

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The techniques for testing digital television systems which incorporate signal processing for the purpose of reducing the number of bits which need to be transmitted to define a video frame are , at this time, poorly defined. In general, purely subjective test procedures have been used to date. It would be desireable to produce a set of quantitative data which correlates directly with a set of qualitative data. The qualitative data will serve as the initial criteri for the evaluation of codec performance. Analysis of the correlation between the quantitative and qualitative data will then permit the development of a set of quantitative tests whose results will serve as a non-subjective standard for the future evaluation of codecs.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dennis Bodson	22b. TELEPHONE (Include Area Code) 202-692-2124 22c. OFFICE SYMBOL NCS-TS

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

NCS TECHNICAL INFORMATION BULLETIN 87-7

DEVELOPMENT OF AN OBJECTIVE STANDARDIZED TEST FOR VIDEO TELECONFERENCING

PROJECT OFFICER

Dennis Bodson

DENNIS BODSON
Senior Electronics Engineer
Office of NCS Technology
and Standards

APPROVED FOR PUBLICATION:

Dennis Bodson

DENNIS BODSON
Assistant Manager
Office of NCS Technology
and Standards

FOREWORD

Among the responsibilities assigned to the Office of the Manager, National Communications System, is the management of the Federal Telecommunication Standards Program. Under this program, the NCS, with the assistance of the Federal Telecommunication Standards Committee identifies, develops, and coordinates proposed Federal Standards which either contribute to the interoperability of functionally similar Federal telecommunication systems or to the achievement of a compatible and efficient interface between computer and telecommunication systems. In developing and coordinating these standards, a considerable amount of effort is expended in initiating and pursuing joint standards development efforts with appropriate technical committees of the Electronics Industries Association, the American National Standards Institute, the International Organization for Standardization, and the International Telegraph and Telephone Consultative Committee of the International Telecommunication Union. This Technical Information Bulletin presents an overview of an effort which is contributing to the development of compatible Federal, national, and international standards in the area of Video Teleconferencing. It has been prepared to inform interested Federal activities of the progress of these efforts. Any comments, inputs or statements of requirements which could assist in the advancement of this work are welcome and should be addressed to:

Office of the Manager
National Communications System
ATTN: NCS-TS
Washington, DC 20305-2010

DEVELOPMENT OF AN OBJECTIVE
STANDARDIZED TEST FOR
VIDEO TELECONFERENCING

FINAL REPORT

February, 1987

SUBMITTED TO:

NATIONAL COMMUNICATIONS SYSTEM

Office of Technology and Standards

Washington, DC 20305

Contracting Agency:

DEFENSE COMMUNICATIONS AGENCY

Contract No. DCA100-83-C-0047
Modification P00009
Task No. 1

D E L T A I N F O R M A T I O N S Y S T E M S, I N C.
Horsham Business Center, Bldg. 3
300 Welsh Road
Horsham, PA 19044

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION AND SUMMARY

This document summarizes work performed by Delta Information Systems, Inc., for the Office of Technology and Standards of the National Communications System, an organization of the U. S. Government, headed by National Communications System Assistant Manager Dennis Bodson. Mr. Bodson is responsible for the management of the Federal Telecommunications Standards Program, which develops telecommunications standards, the use of which is mandatory for all Federal agencies. The purpose of this study, performed under Task number 1, Modification P00009 of Contract number DCA100-83-C-0047, was to determine the feasibility of measuring image quality of video teleconferencing systems using objective rather than subjective procedures.

The techniques for testing digital television systems which incorporate signal processing for the purpose of reducing the number of bits which need to be transmitted to define a video frame are, at this time, poorly defined. In general, purely subjective test procedures have been used to date. It would be desireable to produce a set of quantitative data which correlates directly with a set of qualitative data. The qualitative data will serve as the initial criteria for the evaluation of codec performance. Analysis of the correlation between the quantitative and qualitative data will then permit the development of a set of quantitative tests whose results will serve as a non-subjective standard for the future evaluation of codecs.

This study is an important first step towards the very ambitious goal of establishing objective test methods for digital video codecs. A first effort of this type cannot be expected to immediately accomplish all stated objectives. However, it can be considered successful if it provides good understanding of the applicable criteria and inherent problems, and clearly points the way towards future efforts which will fully accomplish the stated objectives. This report will show that this goal has been achieved.

Section 2 briefly reviews the previously performed subjective tests and provides some additional analysis to convert the results to a format which is more useful for this study. Objective tests are described and analyzed in Section 3. It contains the results of test tape and direct measurements and includes a first attempt at objectively evaluating motion performance. In Section 4 the subjective and objective measurement results are translated into a common format and checked for correlation. Section 5 briefly summarizes the program and makes recommendations for future efforts.

2.0 REVIEW OF SUBJECTIVE TESTS

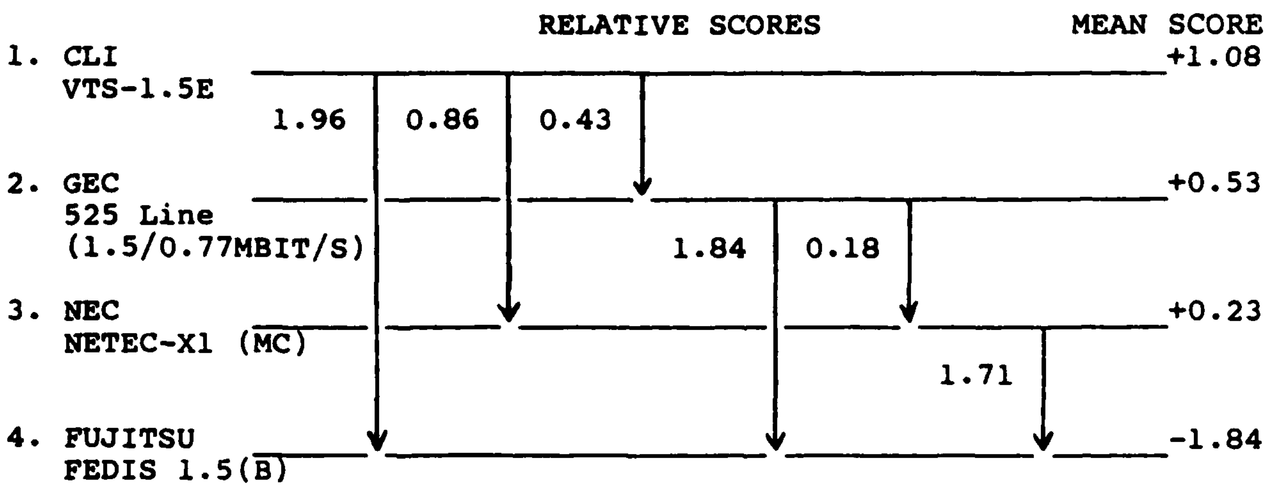
2.1 Analysis of Past Results

During 1984 and 1985, DIS performed extensive subjective evaluations of four video codec models available at that time and operating at 1.544 Mbps. These tests are documented in the Final Report entitled "Test and Evaluation of Teleconferencing Video Codecs Transmitting at 1.5 Mbps" which was submitted to the National Communications System on August 23, 1985. This report summarizes the test results on Table 4-7 which is repeated here for reference as Table 2-1. The score comparison chart shows graphically both the comparative scores of each codec pair and the resulting mean values. The chart is not to scale and the various values cannot add up because they represent means derived in different steps from subjective test scores. However, the results are consistent and thus produce a high confidence in their validity.

The most obvious impairment of codec performance is the rendition of motion which influences the subjective evaluation most heavily. While most analog performance parameters can be measured on a codec without difficulty there is as yet no available methodology for objective measurements of motion performance. It therefore cannot be expected to find much correlation between the subjective test results shown on Table 2-1 and objective analog measurements. Consequently, the content of the DIS codec test tape was reviewed and most sequences put

CODEC NO.	CODEC NO. 1	2	3	4	SUM	MEAN SCORE	RANK	MANUFACTURER
1	-	1.84	-0.43	0.18	1.59	+0.53	2	GEC
2	-1.84	-	-1.96	-1.71	-5.51	-1.84	4	FUJITSU
3	0.43	1.96	-	0.86	3.25	+1.081	1	CLI
4	-0.18	1.71	-0.86	-	0.67	+0.23	3	NEC

RANKING MATRIX



SCORE COMPARISON

TABLE 2-1 SUBJECTIVE CODEC RANKING

OVERALL

into one of three categories, depending on their motion content. These categories are:

1. Still graphics and slow motion
2. Lively motion
3. Camera zooming

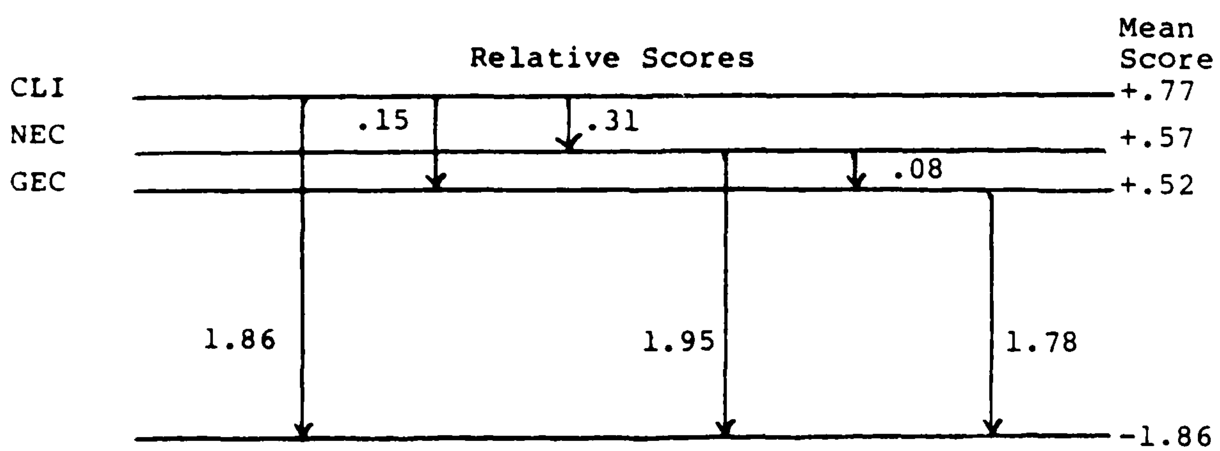
Comparative scores for each codec pair, mean values and codec rankings were computed separately for each category, following the steps previously used in preparing Table 2-1. The results are shown on Tables 2-2, 2-3, and 2-4. The relative rankings are essentially just as consistent as for the full test tape but show some significant differences. In category 1 the scores of CLI, NEC and GEC are very close, almost within the expected margin of error. Category 2 shows values somewhat similar to the full test tape while for Category 3 the spread becomes much larger. This is consistent with the fact that camera zooming stresses the motion capability of a codec most severely.

The slight inconsistencies between relative and mean scores on Tables 2-2 and 2-4 are not errors but inevitable small ambiguities caused by the inherent imperfection of subjective testing. The practical interpretation of the results is that in these cases the difference between GEC and NEC is within the expected margin of error.

In all categories, the score of Fujitsu remains very low. It became obvious that the main problem of this codec is its motion capability, and this may also affect its score even in Category 1. All sequences are switched at start and end, and

	GEC	FUJ	CLI	NEC	SUM	MEAN SCORE	RANK
GEC	-----	+1.78	- .15	- .08	+1.55	+ .52	3
FUJ	-1.78	-----	-1.86	-1.95	-5.59	-1.86	4
CLI	+ .15	+1.86	-----	+ .31	+2.32	+ .77	1
NEC	+ .08	+1.95	- .31	-----	+1.72	+ .57	2

RANKING MATRIX



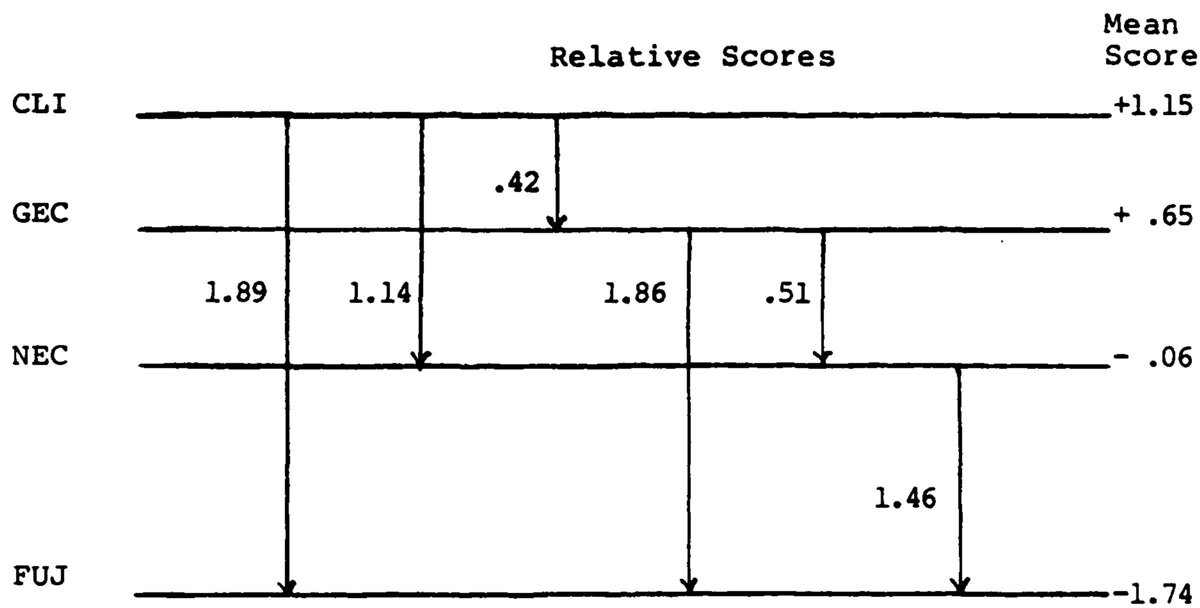
SCORE COMPARISON

Table 2-2

Subjective Codec Ranking
Graphics and Slow Motion

	GEC	FUJ	CLI	NEC	SUM	MEAN SCORE	RANK
GEC	-----	+1.86	- .42	+ .51	+1.95	+ .65	2
FUJ	-1.86	-----	-1.89	-1.46	-5.21	-1.74	4
CLI	+ .42	+1.89	-----	+1.14	+3.45	+1.15	1
NEC	- .51	+1.46	-1.14	-----	- .10	- .06	3

RANKING MATRIX

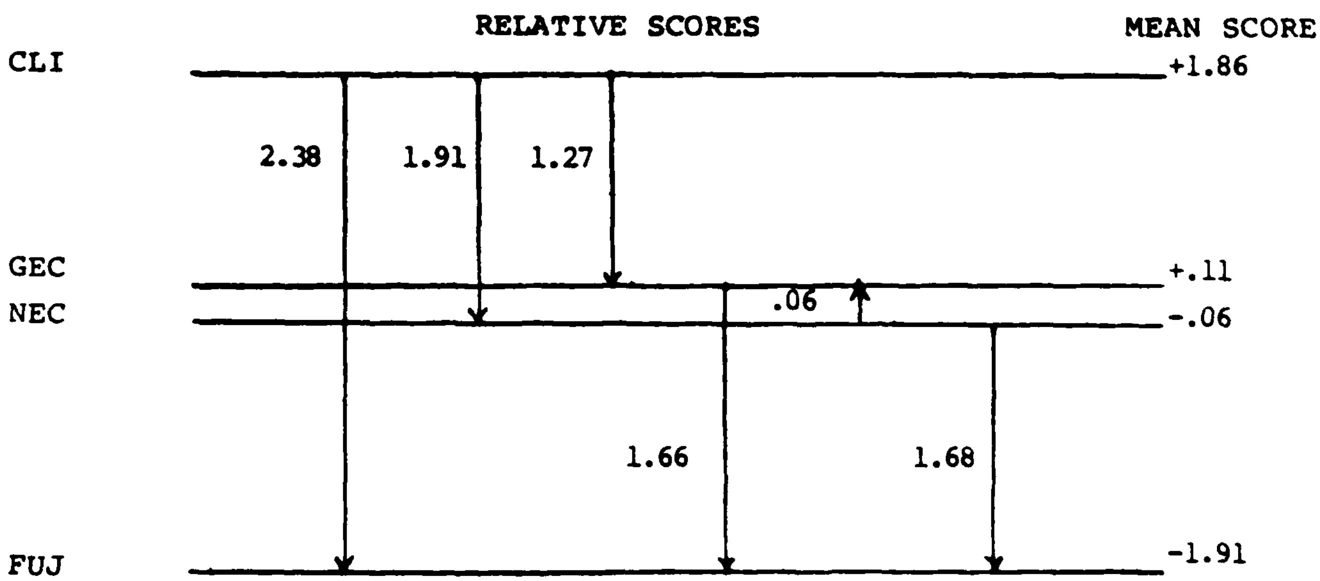


SCORE COMPARISON

Table 2-3
Subjective Codec Ranking
Lively Motion

	GEC	FUJ	CLI	NEC	SUM	MEAN SCORE	RANK
GEC	-----	+1.66	-1.27	- .06	+ .33	+ .11	2
FUJ	-1.66	-----	-2.38	-1.68	-5.72	-1.91	4
CLI	+1.27	+2.38	-----	+1.91	+5.56	+1.86	1
NEC	+ .06	+1.68	-1.91	-----	- .17	- .06	3

RANKING MATRIX



SCORE COMPARISON

Table 2-4

Subjective Codec Ranking
Zoom

many contain a further switch between two different pictures. Any switch represents a kind of motion, and on the Fujitsu codec a switch is reproduced as a vertical wipe. This may have been annoying enough to cause a low score by most evaluators even if the still picture itself is no more than slightly impaired.

Another factor affecting subjective evaluation which presently is unlikely to be identified objectively is artifacts which are uniquely caused by the codec algorithm. Any codec shows occasional spurious output phenomena, such as contours, stripes, squares or other patterns triggered generally by certain features of the input picture. The appearance of such artifacts can be very annoying yet very difficult to predict or identify as to their causes.

2.2 Additional Recent Tests

Video codec technology is in a state of rapid development. All the models on which the 1984 tests were performed are at best obsolescent at this time. The new models are generally interoperable with the older versions. They feature mainly a selection of several lower data rates in addition to improvements in the coding algorithms and operating convenience. The CLI VTS-1.5E has been replaced by the "Rembrandt", and the NEC NETEC-X1 by the NETEC-XV. GEC has no published specific new model designation. Fujitsu showed a simulation of an improved codec in 1984, the design of which has evidently been completed but so far this new model has not yet been made available for independent tests.

During May 1986 DIS performed a large number of subjective codec tests for INTELSAT. Objective tests were considered but could not be implemented because of time limitations. These tests covered a wide range of performance and data rates. In the area of full motion codecs three models were available, namely CLI Rembrandt, NEC NETEC-XV, and Philips VCD-2M. The latter unit follows the European COST-211 standard and is interoperable with the GEC codec. The equipments under test operated at data rates from 384 Kbps to 2.048 Mbps.

A direct comparison between the 1984 and 1986 tests cannot be made because the INTELSAT tests had a different objective. Their purpose was to evaluate the usability of codecs at various data rates for specific typical applications of digital TV. No codec was to be rated individually, and no comparison of the performance of specific codecs was to be made. The test tape consisted mainly of selected scenes from the tape used for the 1984 tests, but re-arranged and edited to meet the INTELSAT requirements.

The test results show that all three codecs are basically acceptable for the selected typical applications over their full range of data rates. As expected, performance improves at higher data rates. Without making an actual comparison, all experienced observers agreed that the new codec models performed better than their predecessors. The differences between models tend to become smaller but remain definitely noticeable. Motion performance continues to be the most important factor in assessing codec quality. Thus, even without a quantitative comparison, the new tests increase confidence in the validity of the previously obtained data.

3.0 OBJECTIVE TESTS

3.1 Purpose

Subjective tests are awkward and time consuming in terms of both execution and evaluation. A large amount of data is necessary to achieve confidence in the results. Yet, the quality and usefulness of a picture should ultimately be judged by the viewer. Analog broadcast TV went through many years of subjective evaluations before it was possible to establish correlation with objective parameters which can be readily measured. There now exist specifications and standards for most TV applications which give performance limits of all pertinent parameters known to determine picture quality.

So far no meaningful objective tests for codecs have been developed. Though many conventional analog tests can be readily performed on a codec, it has not been determined how meaningful these results are and how they relate to a subjective evaluation. Correlation of these two types of tests would greatly facilitate further developments in the digital TV area. Some examples are as follows:

- o Optimization of the parameters of a specific coding algorithm.
- o Comparison of the performance of different coding algorithms.
- o Performance monitoring of TV transmission systems

containing one or several codecs. Limits of acceptable system performance can subsequently be established.

3.2 Parameter Selection

There are three main documents specifying analog TV parameters. They are EIA RS-170A, EIA RS-250B, and NTC Report No. 7. RS-170A gives the basic specifications of the signal waveform. RS-250B and NTC-7 are similar in content and cover the performance parameters likely to be affected by signal processing and transmission. Both documents also give suggested measurement methods and test signals.

The codec encoder processes the analog signal in a radical fashion so that the format of the transmitted compressed signal bears no resemblance to the incoming signal. The decoder reconstitutes the analog signal which means that signal waveforms and timing are generated there and not directly influenced by either the incoming digital signal or the encoding/decoding and transmission processes. Therefore, compliance with the waveform parameters of RS-170A is not dependent on the encoding algorithm and thus not a high priority item for objective testing. On the other hand, most of the parameters specified in RS-250B and/or NTC-7 may be affected by the encoding algorithm and therefore should be considered for an objective test program.

There are other factors unique to codecs which affect some of the objective test parameters. Since codecs normally "clip" the transmitted picture by reducing both width and height, both horizontal and vertical blanking will be intentionally wider than specified in RS-170A. Parameters which are mainly affected by certain factors typical of analog transmission become largely irrelevant in a digital transmission system. Non-linear transfer characteristics and dynamic gain distortions are often caused by limitations in FM detectors and low frequency response. Therefore, measurements of dynamic gain, long time waveform distortion (bounce), and use of average picture levels (APL) other than 50% in differential gain and phase measurements become unnecessary. Transmission noise of all types is highly unlikely to affect the received picture because the decoder is tolerant to error rates up to 10^{-6} before forward error correction. The only noise to be considered is a sum of quantizing noise and contributions from power supplies and other portions of the circuit. This noise level is inherently low. The output level of the re-constituted signal (often called insertion gain), once properly set, is most likely to stay constant. Field time waveform distortion caused by low frequency response limitations will be low and constant. Therefore, the number of important parameters for objective testing can be considerably reduced.

3.3 Measurements

3.3.1 Video Tape Tests

The test tape prepared for the 1984 comparative codec tests consists of two parts. The first part contains the scenes for strictly subjective evaluation which were used for the tests described in paragraph 2.1. The second part contains a variety of test signals to be used partly for fully objective measurements and partly for viewing by video experts with the anticipation that basically subjective but possibly semi-objective results may be obtained. Both parts were processed through the four codecs under test and the outputs recorded on 1" video tape.

Table 3-1 gives the scenario for the test signal portion of the test tape. The selection of the various signals was based on both established practice and reasonable expectations of what may be accomplished. Sequences 1 to 14 contain conventional signals largely used for evaluation and objective measurements of analog TV signals. Sequences 15 to 18 contain artificial controlled motion and were designed to implement initial attempts to develop a methodology for objective measurement or semi-objective evaluation of motion performance. Specifically, sequences 15 and 16 contain switching between two radically different pictures for the purpose of simulating fast motion.

The very straightforward test arrangement is shown on Figure 3-1. The test tape to be analyzed is played in the 1" tape recorder which is equipped with the appropriate time base corrector and allows frame-by-frame manual advance. The signal waveform is shown directly on a waveform monitor and a vector display is presented on a vectorscope when needed. The picture is also viewed on a high quality monitor. An oscilloscope camera,

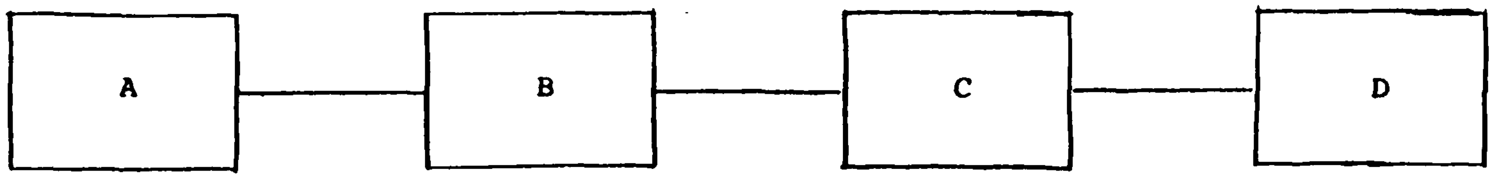
TABLE 3-1

TEST SIGNAL TAPE SCENARIO

Sequence No.	Description	Purpose
1	Color Bar Chart	Level check, vector accuracy measurement
2	RETMA Resolution Chart	Resolution check
3	Gray Scale Chart	Subjective check of appearance of flat gray areas
4	Composite Test Signal - consisting of modulated 12.5T sine square pulse, 2T pulse, vertical white bar	Chrominance-to luminance gain and delay inequality, line time and short time waveform distortion
5	Video Sweep (with markers)	Amplitude vs. frequency response, filter parameters,
6	5 Step staircase, APL=50%	Luminance nonlinearity
7	5 Step staircase, APL=90%	Dynamic gain of picture signal
8	5 Step staircase, APL=10%	and sync signal
9	5 Step modulated staircase, APL= 50%	Differential gain
10	5 Step modulated staircase, APL= 90%	Differential phase
11	5 Step modulated staircase, APL= 10%	
12	Ramp	Luminance sampling precision Quantization noise, Contouring
13	Modulated Ramp	Chrominance sampling precision Differential gain and phase
14	3 Level Chroma	Chrominance-to-luminance intermodulation Chrominance nonlinear gain and phase

TABLE 3-1 (cont'd)

Sequence No.	Description	Purpose
15	Switch between white window & black field, 3 times 10 seconds each.	Objective or subjective motion measurement.
16	Switch between yellow & blue field, 3 times 10 seconds each.	Objective or subjective motion response evaluation.
17	Vertical black bar moving across white field, both directions, various speeds	Subjective evaluation of artificial motion, compared to natural motion
18	Black diamond shaped outline on white field, moving out from and in toward center, various speeds	Subjective evaluation of artificial motion, compared to natural motion



LEGEND

A. 1" Video Tape Recorder Sony BVH 1100A with Time Base Corrector Sony BVT 2000

B. TV Wave Form Monitor Textronix 1480R

C. Vectorscope Tektronix 520A

C. NTSC Color Monitor Barco CTVM 2.51

Figure 3-1 Test Tape Analysis Arrangement

Tektronix C-4, is used to photograph selected waveforms, vector displays and monitor pictures.

The results of the tests did not come up to expectations. When measuring the test tape before processing through a codec it was found that some of the signals contained a considerable amount of distortions. It cannot be determined after the fact whether this was due to imperfections in the test signal generators or to a problem in the taping process. It will be shown subsequently that adjustments of the measurements were made to achieve meaningful results.

One important result of the measurements was that not all test signals are suitable for use with digital video codecs. Though the encoding algorithms differ between codecs, they all have some or all of the features of bandwidth limitation, horizontal and vertical sampling and sub-sampling, and interpolation. These factors introduce distortions such as aliasing and full or partial suppression of some test signal portions. Some of these distortions can be recognized and discarded but the tests made it obvious that some modifications of test signals are necessary to make meaningful objective measurements on digital TV codecs.

Most values were read from the waveform monitor or vectorscope screens using the standard graticules. In some cases waveforms were photographed with the oscilloscope camera. They were color bar chart, differentiated unmodulated ramp, modulated ramp through a 3.58 MHz band pass filter, and video sweep at vertical (field) scanning rate. In addition, the picture of the video sweep was photographed on the monitor screen. These photographs are shown on Figures 3-2 to 3-6.

The photographs of the color bar chart vector display can be used to determine amplitude and phase errors. The video sweep display shows the frequency response but great care must be taken to properly interpret the pattern because the complex processing of the codec produces aliasing and other spurious patterns. The photographs of the monitor screen help in identifying the meaning of the various portions of the sweep display. They show that signal amplitudes appearing on the waveform monitor at and above 3 MHz consist of lower frequencies produced by aliasing and similar phenomena cause by sampling, interpolation and filtering and thus do not depict a real response. This cannot be readily identified on the waveform monitor. Generally only the first part of the pattern with an envelope decreasing from a high value at a low frequency to zero is a true representation of the codec response. The area of this envelope gives a measure of the response.

The unmodulated and modulated ramp signals were expected to yield measures of quantizing noise and sampling accuracy. However, these signals were so much contaminated by spurious components to make it impossible to derive any meaningful quantitative data from these displays.

Table 3-2 lists the results that were obtained. It shows the test parameters, test signals, and the results of the

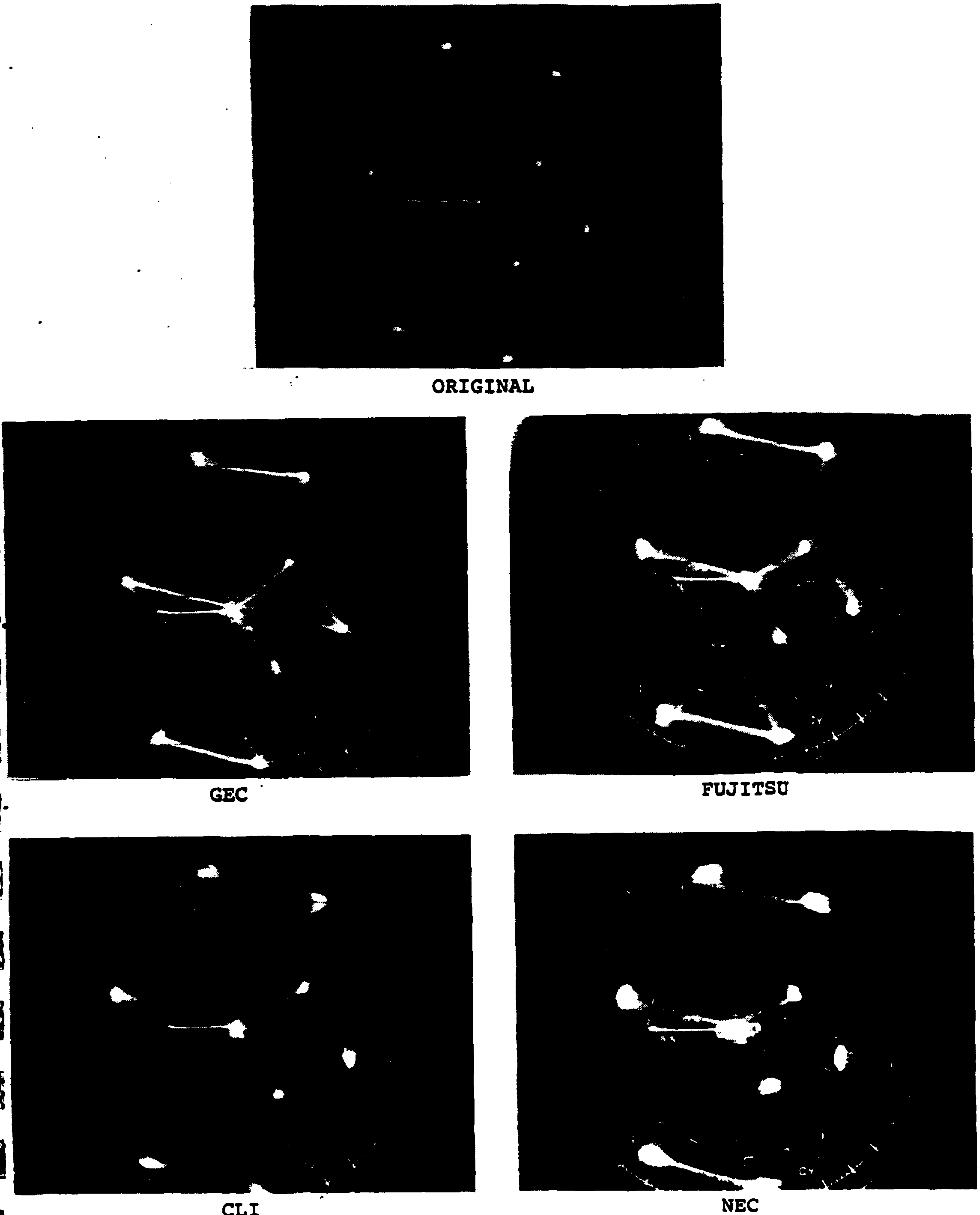
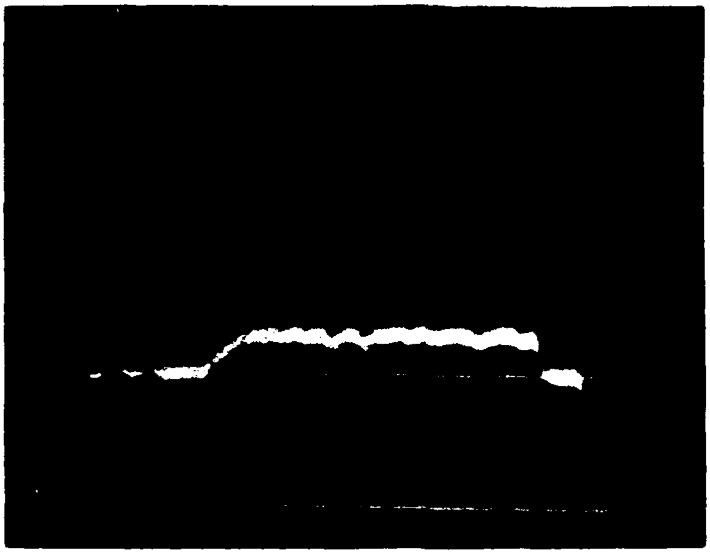
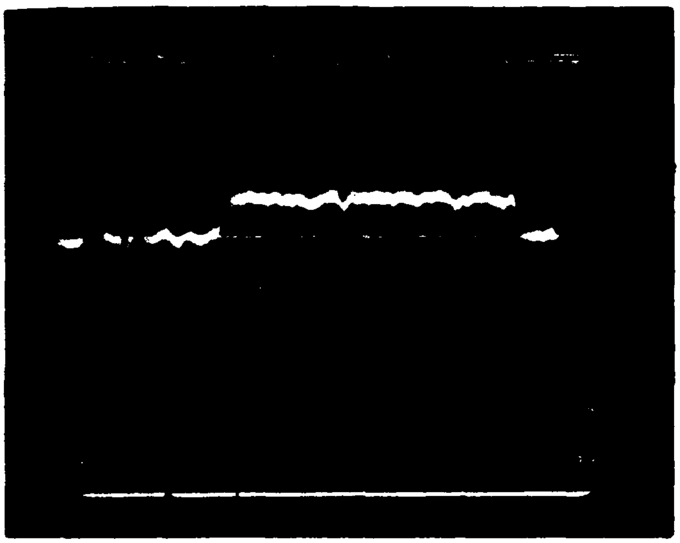


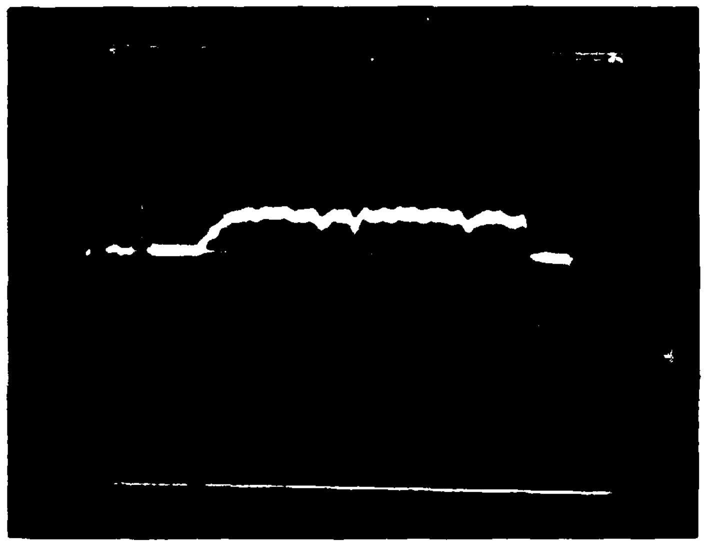
Figure 3-2 Color Bar Chart Patterns on Vectorscope



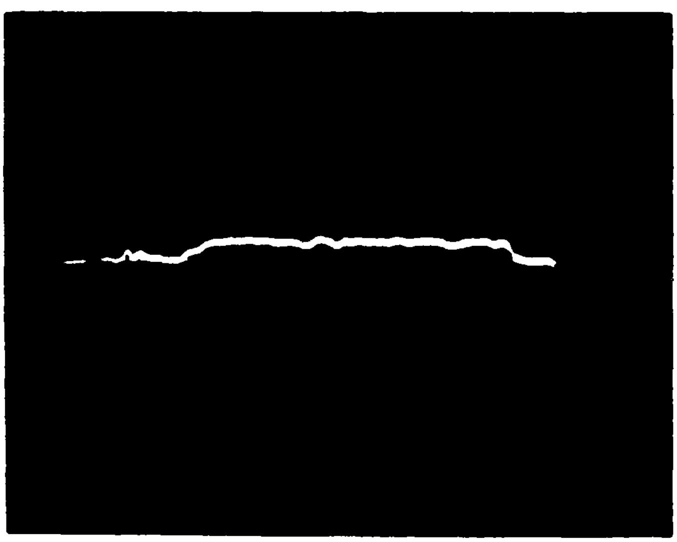
ORIGINAL



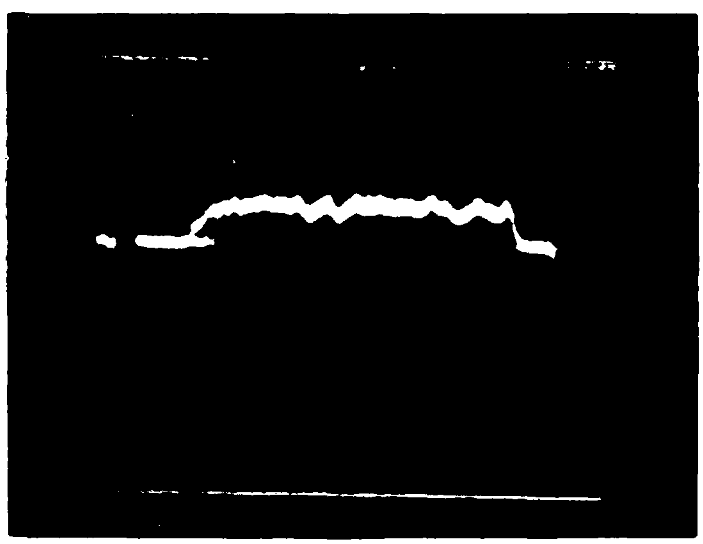
GEC



FUJITSU

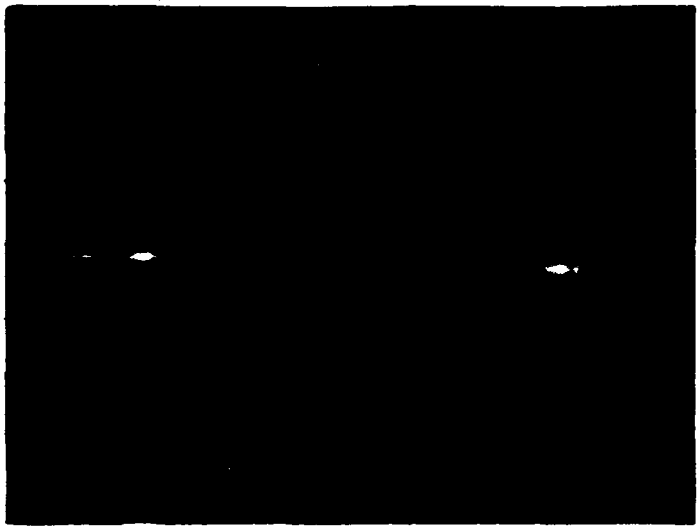


CLI

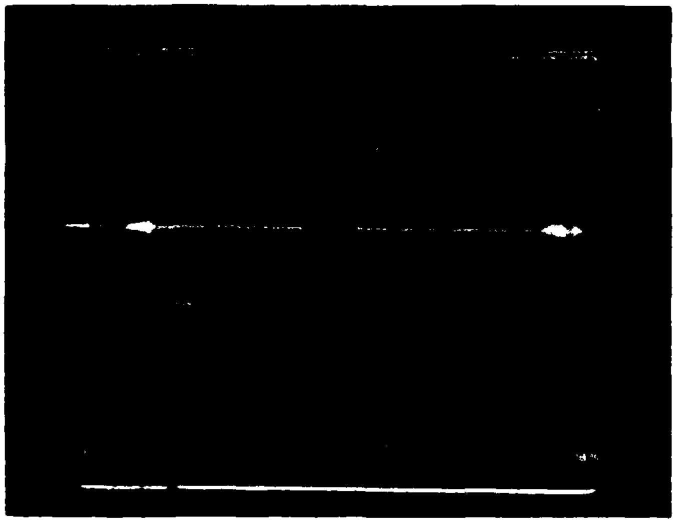


NEC

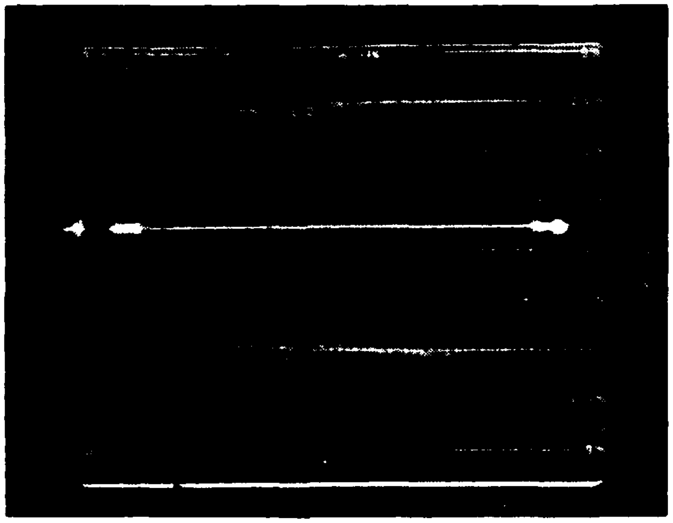
Figure 3-3 Differentiated Unmodulated Ramp on Waveform Monitor



ORIGINAL

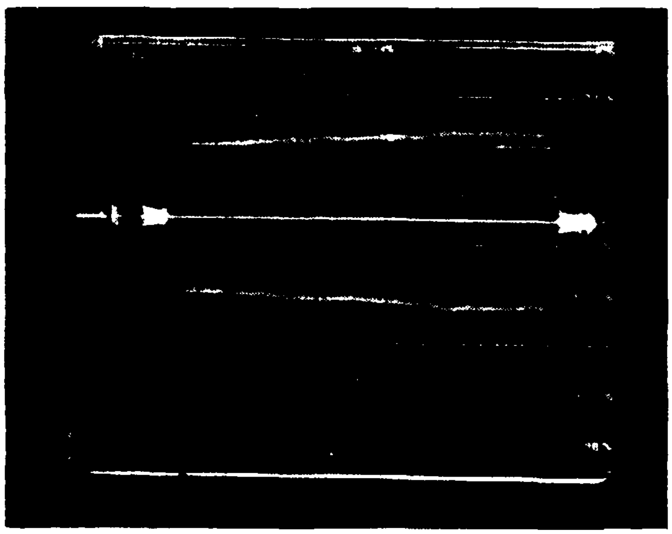


GEC



FUJITSU

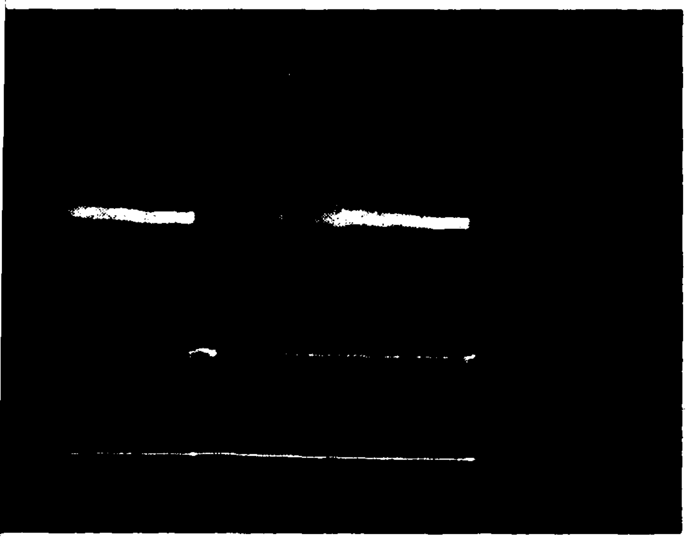
CLI



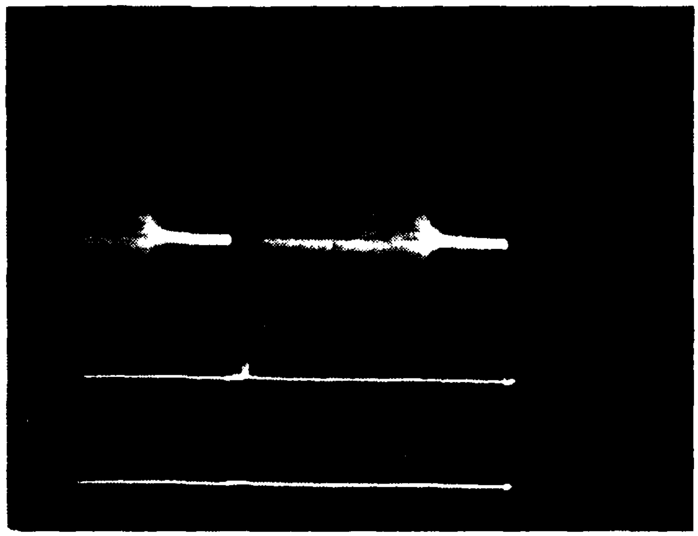
NEC

Figure 3-4 Modulated Ramp through Band Pass Filter on Waveform Monitor

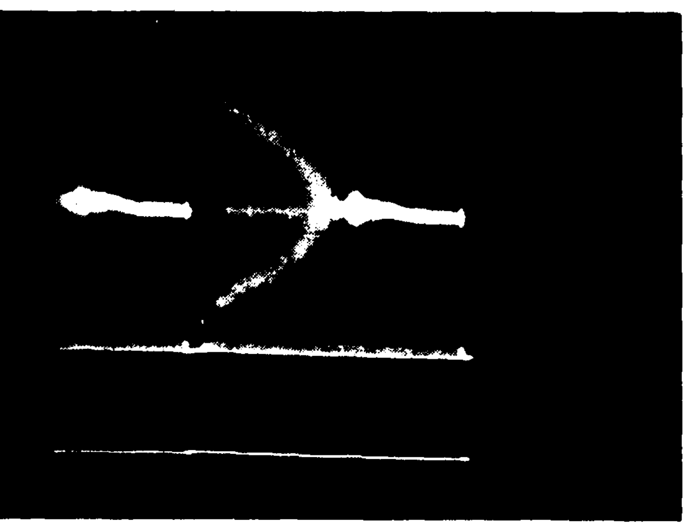
ORIGINAL



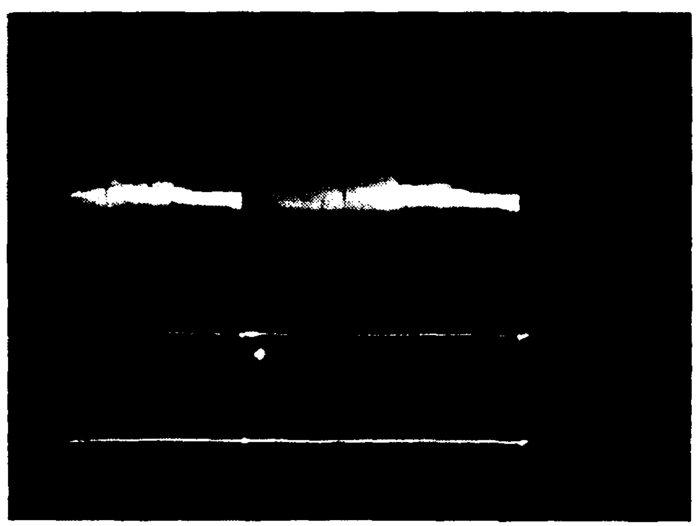
GEC



FUJITSU

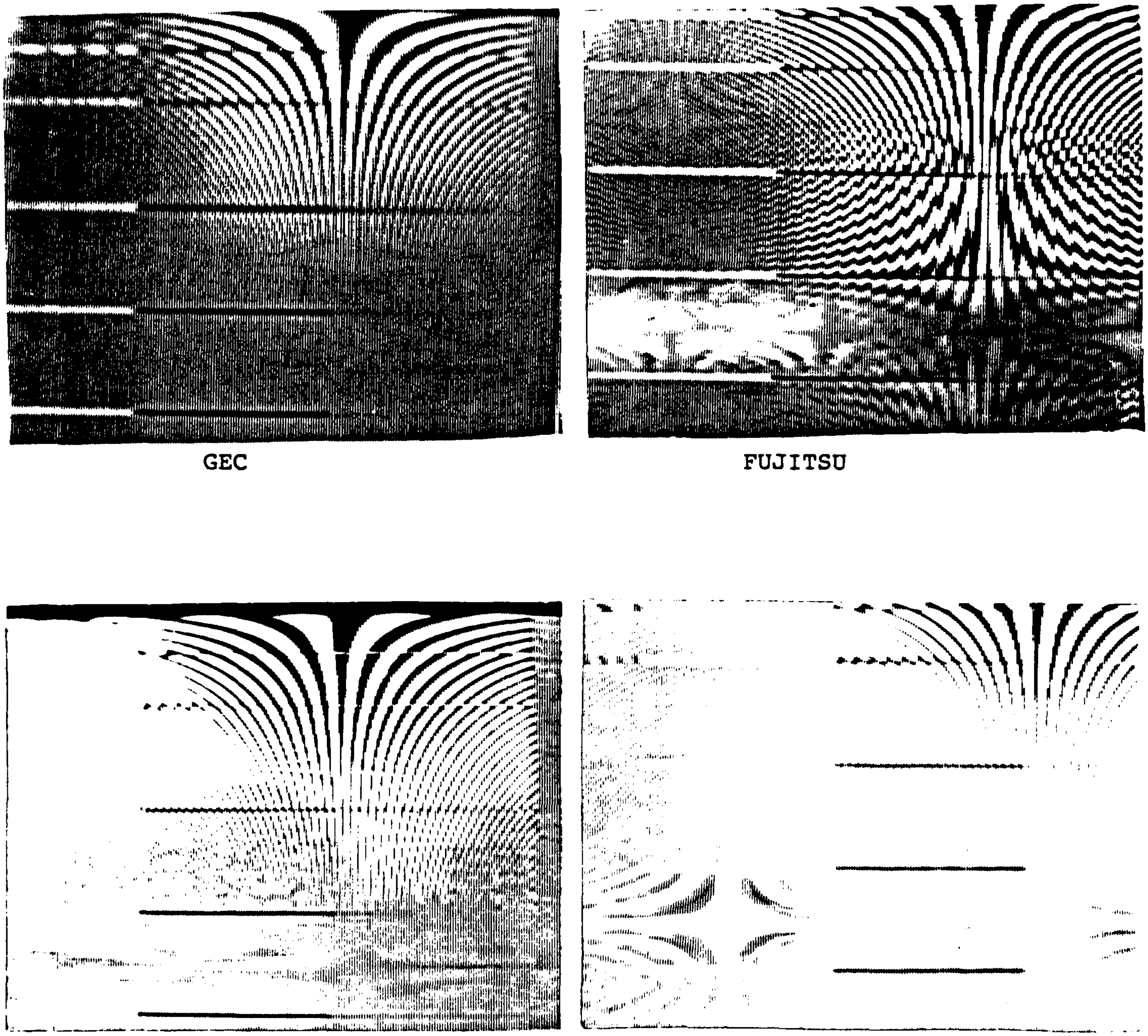


CLI



NEC

Figure 3-5 Field Rate Video Sweep on Waveform Monitor



GEC FUJITSU

CLI NEC

Figure 3-6 Field Rate Video Sweep on Picture Monitor

TABLE 3-2 VIDEO TAPE TESTS

NO	PARAMETER		UNIT	TEST SIGNAL	ORIGINAL TAPE	MEASUREMENTS GEC	FUJ	CLI	NEC	STANDARDS RS-250B	NTC-7	NOTES
1	FREQUENCY RESPONSE		FIGURE OF MERIT	VIDEO SWEEP (VERTICAL RATE)	1000	240	110	200	220	±1.2DB (TO 4.2MHZ)	+3/-5IRE (TO 4.2MHZ)	NTC-7 STANDARD IS BASED ON MULTIBURST TEST SIGNAL.
2	RESOLUTION	HOR	LINES	RESOLUTION CHRT	450	210	<200	250	250	-----	-----	
		VERT			400	200	<200	250	300			
3	CHROM/LUM GAIN INEQUALITY		IRE	12.5T MOD. PULSE	-3	-15	0	-10	-10	±7	±3	
4	CHROM/LUM DELAY INEQUALITY		NSEC	12.5T MOD. PULSE	0	312	165	171	0	±50	±75	
5	LINE TIME WAVEFORM DISTORTION		IRE	LINE BAR	1	1	1	1	1	2	4	
6	SHORT TIME WAVEFORM DISTORTION	2T PULSE	IRE	2T PULSE	+5	-13	-5	-14	-12	---	±6	MEASURED DATA INCOMPLETE.
		OVERSHOOT	IRE			---	+4/-7	±10	---	7	10	
		K FACTOR	%		±3	---	---	+1/-2	---	---	---	
7	LUMINANCE NONLINEARITY	APL=10%	%	UNMODULATED STAIRCASE	20	---	30	16	20	10	10	NO DATA FOR GEC BECAUSE OF UN-EXPLAINABLE HALF AMPLITUDE OF LOWEST STAIRCASE STEP.
		APL=50%			20	---	30	20	24	10	10	
		APL=90%			21	---	30	24	24	10	10	
8	DIFFERENTIAL GAIN	APL=10%	%	MODULATED STAIRCASE	16	---	---	---	16	10	15	MEASURED DATA INCOMPLETE.
		APL=50%			15	17	16	20	14	10	15	
		APL=90%			18	---	---	---	17	10	15	
9	DIFFERENTIAL PHASE	APL=10%	DEGREES	MODULATED STAIRCASE	12	---	---	---	---	3	5	MEASURED DATA INCOMPLETE.
		APL=50%			13	13	11	13	16	3	5	
		APL=90%			7	---	---	---	---	3	5	
10	CHROM/LUM INTERMODULATION		%	THREE LEVEL CHROMA	2	2	2	2	1	4	3	
11	CHROMINANCE NON-LINEAR GAIN		%	THREE LEVEL CHROMA	+1/+2	0/+3	+1/+3	0/+3	+1/+1	5	±2/±8	THE TWO VALUES ARE THE AMPLITUDE ERRORS OF THE 20 AND 80IRE LEVELS, WITH THE 40IRE LEVEL SET ACCURATELY.

12	CHROMINANCE NON-LINEAR PHASE		DEGREES	THREE LEVEL CHROMA	3	1	2	3	7	5	5	
		YELLOW AMPL.ERROR	%		0	0	+2	+8	+2	---	---	
		PHASE ERROR	DEGREES		+2	+5	+5	+2	+2	---	---	
		CYAN AMPL.ERROR	%		+2	-2	0	+7	+2	---	---	
		PHASE ERROR	DEGREES		+2	+2	+4	+7	+5	---	---	
		GREEN AMPL.ERROR	%		+2	0	+4	+6	+2	---	---	RED BAR AND BURST SET ACCURATELY.
13	VECTOR ACCURACY	PHASE ERROR	DEGREES	COLOR BAR CHART	+2	+2	+4	+6	+6	---	---	
		MAGENTA AMPL.ERROR	%		0	+3	0	+2	0	---	---	
		PHASE ERROR	DEGREES		0	-2	0	+2	+2	---	---	
		RED AMPL.ERROR	%		0	0	0	0	0	---	---	
		PHASE ERROR	DEGREES		0	0	0	0	0	---	---	
		BLUE AMPL.ERROR	%		0	-2	-3	0	-3	---	---	
		PHASE ERROR	DEGREES		0	-4	+2	+2	+2	---	---	

original test tape and through the four codecs under test. The difference between the original and processed tapes yields the corrected values which will be used in further analysis.

Several explanatory remarks to Table 3-2 are necessary. The frequency response patterns shown on Figure 3-2 are difficult to analyze because of aliasing and other distortions introduced by the codec algorithm. A numerical result was obtained as follows:

$$\text{FIGURE OF MERIT} = 1000 \int_0^Z Yfdf \Big/ \int_0^{4.2} Xfdf$$

where X = Sweep amplitude at codec input

Y = Sweep amplitude at codec output

Z = Frequency of zero response at codec output as seen on waveform monitor

The integration is performed on a point-by-point basis.

This method of computation takes variations in the input signal into account. The short time waveform distortion measurements are affected by the limited frequency response; in addition, an unexplained contamination of the signal with a low amplitude color subcarrier made measurement of overshoots impossible in several cases. Incomplete measurements of differential gain and phase are due to the fact that at an APL of 10% and 90% color subcarrier exists only on every fifth line. This is not compatible with chrominance vertical subsampling and subsequent interpolation in the codec resulting in a very low subcarrier amplitude and therefore a high noise level which makes meaningful measurements impossible.

3.3.2 Direct Measurements

As part of another program, DIS made complete objective measurements on two CLI Rembrandt codecs. Though this is a more recent and improved model, the changes in the signal processing portion are small and the encoding algorithm is identical with the one in the earlier VTS-1.5E. Therefore, the test results are an applicable input to check and verify some of the other results of this program.

A block diagram of the test setup is shown on Figure 3-7. The tests covered many more parameters than required for codec testing and still did not come close to utilizing the total capability of the equipment. A minor limitation was due to the fact that all test signals were those which are conventionally used by the broadcast industry, meaning that multiburst and chroma pulse were not optimized for the requirements of the codec. Each test encompassed a complete encoder/decoder combination from analog input to analog output and included all elements of internal digital processing.

The most important element in the tests was the TEKTRONIX 1980 ANSWER equipment which allows the collection of large amounts of highly accurate data in a very short time. It requires an external display terminal for entering commands and display of the test results. After the results have been reviewed, they are fed to the printer one page at a time. All parameter measurements are programmed in ANSWER and arranged in

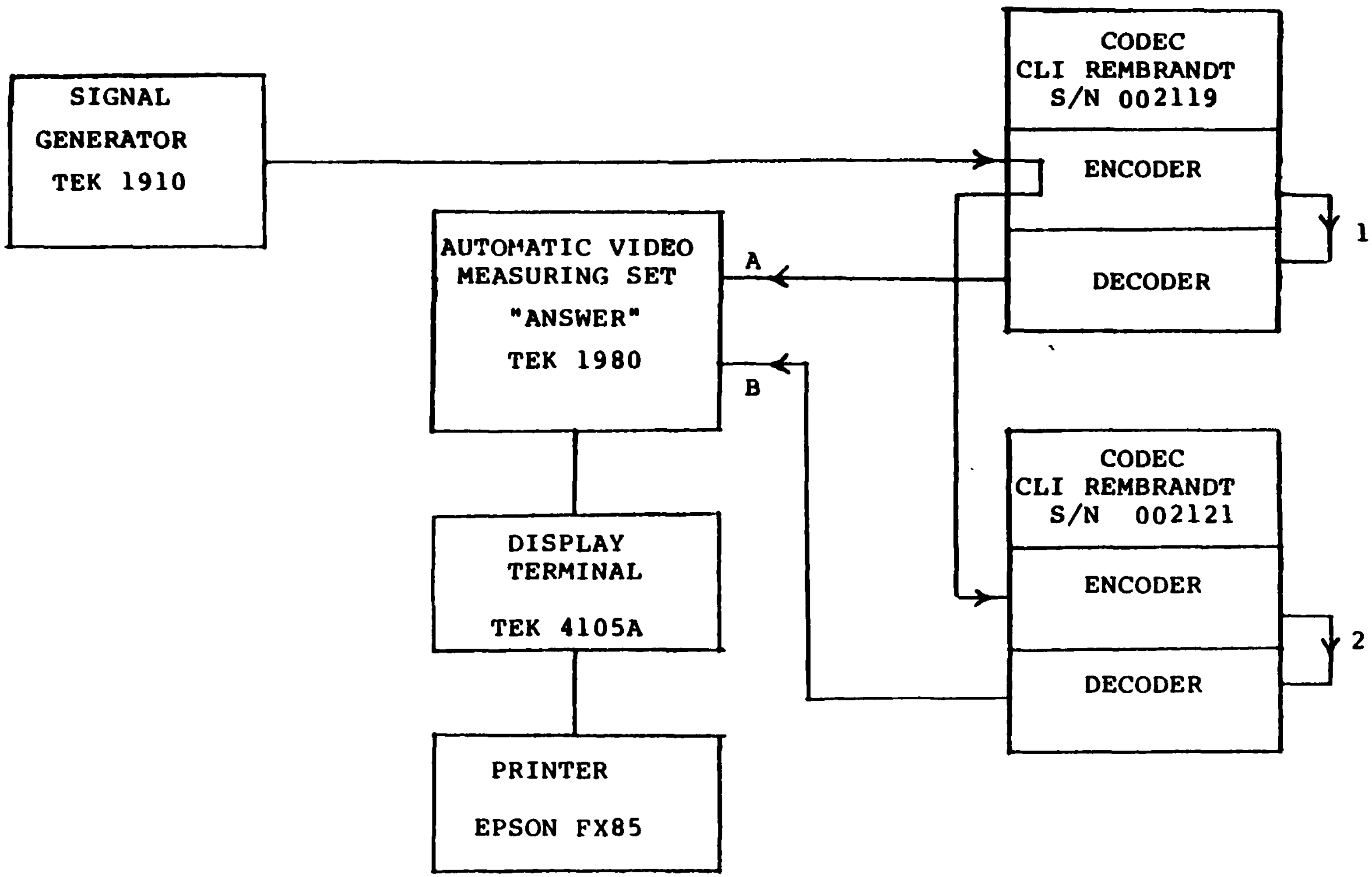


FIGURE 3-7 CODEC TEST SETUP

desirable groupings for ease in commanding. Should other groupings be more convenient, they can easily be programmed.

ANSWER makes all measurements on a single line, utilizing every element of the signal appearing on this line. In the case of the tests described herein, full field test signals were used. In this case, any line during the picture interval can be used for measurement with no effect on the results, and one line was selected arbitrarily. The test signals were selected in the 1910 Signal Generator and the ANSWER Test Set was commanded to measure the parameters which can be handled by each test signal. The results were first viewed on the screen of the display terminal and subsequently printed. Copies of the printouts are shown on Tables 3-3 to 3-6.

A review of the test results shows that, with the exception of field time waveform distortion which cannot be measured on a single line and was not programmed into the available unit of ANSWER, all pertinent TV Signal parameters have been covered. A few words of clarification are needed, mainly because the test equipment was set up for standard broadcast video performance. The amplitude-frequency response is given only as relative amplitudes of the six frequency packets of the multiburst signal which must be compared to their original amplitude of 60%. The packet frequencies are .5, 1, 2, 3, 3.58 and 4.2 MHz which is not compatible with the limitations of the codec and gives only three (3) useful points of measurement. The chrominance pulse has its conventional width of 12.5T instead of the 20T (T = 125 nsec)

```
                     TEKTRONIX VIDEO MEASUREMENTS
12-NOV-85 15:26:22                         VIOLATED LIMITS
                                            LOWER   UPPER


  CHANNEL A  (S/N 002119)  NTSC  15:26:29  APL = 48%
  MEASURING FIELD 1, LINE 65
MULTIBURST FLAG       100.9  IRE                            100 IRE = 714 mV
FCC MB PACKET #1       53.0  % FLAG
FCC MB PACKET #2       50.2  % FLAG
FCC MB PACKET #3       43.7  % FLAG **   45.0    75.0
FCC MB PACKET #4        7.1  % FLAG **   45.0    75.0
FCC MB PACKET #5        1.3  % FLAG **   45.0    75.0
FCC MB PACKET #6         .6  % FLAG *    45.0    75.0

  CHANNEL B  (S/N 002121)  NTSC  15:27:06  APL = 46%
  MEASURING FIELD 1, LINE 65
MULTIBURST FLAG       100.6  IRE                            100 IRE = 714 mV
FCC MB PACKET #1       56.9  % FLAG
FCC MB PACKET #2       54.8  % FLAG
FCC MB PACKET #3       50.4  % FLAG
FCC MB PACKET #4        8.2  % FLAG **   45.0    75.0
FCC MB PACKET #5        1.0  % FLAG **   45.0    75.0
FCC MB PACKET #6         .3  % FLAG *    45.0    75.0
(COMMANDS DONE)
PAGE,MESURA,COMB,MESURB,COMB
PAGE,MESURA,COMB,MESURB,COMB RECEIVED
                     TEKTRONIX VIDEO MEASUREMENTS
12-NOV-85 15:28:27                         VIOLATED LIMITS
                                            LOWER   UPPER


  CHANNEL A  (S/N 002119)  NTSC  15:28:34  APL = 56%
  MEASURING FIELD 1, LINE 65
NTC7 20 IRE CHROMA     20.2  IRE                            (REF 40 IRE CHR)
NTC7 80 IRE CHROMA     75.6  IRE                            (REF 40 IRE CHR)
NTC7 CHR NL PHASE       2.9  DEG
NTC7 CHR-LUM INTMD       .8  IRE                            (REF LUM PED)

  CHANNEL B  (S/N 002121)  NTSC  15:29:06  APL = 54%
  MEASURING FIELD 1, LINE 65
NTC7 20 IRE CHROMA     20.8  IRE                            (REF 40 IRE CHR)
NTC7 80 IRE CHROMA     70.3  IRE       **   71.5    88.5    (REF 40 IRE CHR)
NTC7 CHR NL PHASE       3.7  DEG
NTC7 CHR-LUM INTMD       .8  IRE                            (REF LUM PED)
(COMMANDS DONE)
```

TABLE 3-3

MEASUREMENTS WITH COMBINATION TEST SIGNAL

TEKTRONIX VIDEO MEASUREMENTS

12-NOV-85 17:19:05 VIOLATED LIMITS

CHANNEL A (S/N 002119) NTSC 17:19:12 APL = 61%

MEASURING FIELD 1, LINE 65

				LOWER	UPPER	
BLANKING LEVEL	------	% CARR				
BAR AMPLITUDE	96.3	IRE				100 IRE = 714 mV
SYNC AMPLITUDE	41.7	% BAR				
BLANKING VARIATION	.6	% BAR				
SYNC VARIATION	.7	% BAR				
BURST AMPLITUDE	99.0	% SYNC				
H BLANK 4 IRE	11.91	USEC	**	10.49	11.16	
SYNC WIDTH	4.77	USEC				
SYNC RISETIME	95.0	NSEC				
SYNC FALLTIME	90.0	NSEC				
SYNC-TO-SETUP	10.06	USEC				
FRONT PORCH	1.31	USEC	*	1.4	999.99	
SYNC-TO-BURST-END	7.64	USEC				
BREEZEWAY	.35	USEC	**	.38	999.99	
BURST WIDTH	9.0	CYCLES				
EQUALIZER WIDTH	49.4	% S.W.				
SERRATION WIDTH	4.79	USEC				
V BLANK 4 IRE F1	22.0	LINES	**	18.3	21.1	
V BLANK 4 IRE F2	23.0	LINES	**	18.3	21.1	
LINE TIME DIST	.9	%				
PULSE/BAR RATIO	82.0	%	**	94.0	106.0	
SCH PHASE	-71.0	DEG	*	-45.0	45.0	
CHROMA-LUM DELAY	41.0	NSEC	*	-40.0	40.0	
CHROMA-LUM GAIN	97.8	%				
DIFF GAIN (DG)	4.2	%				
DIFF PHASE (DP)	1.7	DEG				
LUM NL DIST (DY)	9.5	%				
REL BURST GAIN	-7.6	%				
REL BURST PHASE	-1.1	DEG				
2T PULSE RINGING	2.8	% KF				
2T BAR DIST (LD)	1.8	% KF				RINGING
2T BAR DIST (TR)	.8	% KF				RINGING

(COMMANDS DONE)

TABLE 3-4

MEASUREMENTS WITH COMPOSITE TEST SIGNAL

TEKTRONIX VIDEO MEASUREMENTS

12-NOV-85 17:23:09

CHANNEL B (S/N 002121) NTSC 17:23:21 APL = 58%
MEASURING FIELD 1, LINE 65

				VIOLATED LIMITS LOWER	UPPER	
BLANKING LEVEL	------	% CARR				
BAR AMPLITUDE	97.6	IRE				100 IRE = 714 mV
SYNC AMPLITUDE	40.2	% BAR				
BLANKING VARIATION	.5	% BAR				
SYNC VARIATION	.7	% BAR				
BURST AMPLITUDE	101.4	% SYNC				
H BLANK 4 IRE	12.49	USEC	**	10.49	11.16	
SYNC WIDTH	4.75	USEC				
SYNC RISETIME	80.0	NSEC				
SYNC FALLTIME	75.0	NSEC				
SYNC-TO-SETUP	13.47	USEC				
FRONT PORCH	1.74	USEC				
SYNC-TO-BURST-END	7.59	USEC				
BREEZEWAY	.42	USEC				
BURST WIDTH	8.7	CYCLES				
EQUALIZER WIDTH	49.3	% S.W.				
SERRATION WIDTH	4.77	USEC				
V BLANK 4 IRE F1	22.0	LINES	**	18.3	21.1	
V BLANK 4 IRE F2	23.0	LINES	**	18.3	21.1	
LINE TIME DIST	1.2	%				
PULSE/BAR RATIO	84.8	%	**	94.0	106.0	
SCH PHASE	27.5	DEG				
CHROMA-LUM DELAY	-3.0	NSEC				
CHROMA-LUM GAIN	92.6	%				
DIFF GAIN (DG)	4.0	%				
DIFF PHASE (DP)	1.7	DEG				
LUM NL DIST (DY)	18.7	%	*	0.0	10.0	
REL BURST GAIN	-9.9	%				
REL BURST PHASE	-2.6	DEG				
2T PULSE RINGING	2.7	% KF				
2T BAR DIST (LD)	1.9	% KF				RINGING
2T BAR DIST (TR)	3.3	% KF				RINGING

(COMMANDS DONE)

TABLE 3-5

MEASRUEMENTS WITH COMPOSITE TEST SIGNAL

TEKTRONIX VIDEO MEASUREMENTS

12-NOV-85 15:24:49 VIOLATED LIMITS

LOWER UPPER

CHANNEL A (S/N 002119) NTSC 15:24:55 APL = 50%

MEASURING FIELD 1, LINE 65

FCC COLOR BARS:

	AMPL ERROR %	PHASE ERROR DEG	CHR/LUM RATIO % NOM
YEL	13.3	-.3	114.3
CYN	6.0	-3.8	107.3
GRN	10.9	-1.7	112.9
MAG	10.0	-3.3	112.2
RED	3.7	-4.1	105.5
BLU	8.5	.8	112.0

CHANNEL B (S/N 002121) NTSC 15:25:27 APL = 44%

MEASURING FIELD 1, LINE 65

FCC COLOR BARS:

	AMPL ERROR %	PHASE ERROR DEG	CHR/LUM RATIO % NOM
YEL	5.9	2.5	108.2
CYN	-.2	-3.2	102.3
GRN	2.4	1.6	104.4
MAG	5.2	-.7	108.2
RED	-.6	-3.6	100.1
BLU	1.7	4.2	104.4

(COMMANDS DONE)

TABLE 3-6

MEASUREMENTS WITH COLOR BAR CHART

recommended for many codecs. The limits for caution (*) and alarm (**) are set up for broadcast performance and are of no significance for this program.

3.3.3 Measurement Summary

Table 3-7 gives the summary of the measurements described in the two preceding paragraphs. It contains the parameters which are common to Table 3-2 and Tables 3-3 to 3-6. It maintains the format of Table 3-2 and all notes are equally applicable. The values listed under corrected tape measurements are the codec output values minus the measurements on the original tape as listed on Table 3-2. This compensates for the deficiencies in the input signal and isolates the contributions of the codecs. Since in many cases only the absolute values are significant, several minus signs have been dropped. The values listed under direct measurements are the averages of the results obtained on both codecs under test.

The codec output frequency response measurements have already taken the imperfections of the input tape into account and do not require correction. In addition, a figure of merit for the directly measured response had to be computed which was done by using the 6 measured multiburst frequency packet amplitudes given on Table 3-3. This gives the result of 236. However, the values at 3 MHz and above are questionable and probably invalid due to spurious responses. Using only the first three packets yields a figure of merit of 202 which is likely to

TABLE 3-7 MEASUREMENT SUMMARY

NO	PARAMETER		UNIT	CORRECTED TAPE MEASUREMENTS GEC	FUJ	CLI	NEC	DIRECT MEASUREMENT CLI
[illegible]	FREQUENCY RESPONCE		FIGURE OF MERIT	240	110	200	220	236(202)
3	CHROM/LUM GAIN INEQUALITY		IRE	-12	[illegible]	-	-	-5
4	CHROM/LUM DELAY INEQUALITY		nsec	312	163	171	0	19
5	LINE TIME WAVEFORM DISTORTION		IRE	1	1	1	1	1
6	SHORT TIME WAVEFORM DISTORTION	2T PULSE	IRE	-16	-10	-19	-17	-17
		K FACTOR	%	---	---	-2/+1	---	+3
7	LUMINANCE NONLINEARITY (APL=50%)		%	---	10	0	4	14
8	DIFFERENTIAL GAIN (APL=50%)		%	2	1	5	1	4
9	DIFFERENTIAL PHASE (APL=50%)		DEGREES	3	2	[illegible]	3	2
10	CHROM/LUM INTERMODULATION		%	0	0	0	1	1
11	CHROMINANCE NON-LINEAR GAIN		%	-1/+1	0/+1	-1/+1	0/-1	+1/-7
12	CHROMINANCE NON-LINEAR PHASE			2(2)	1(4)	0(3)	4(6)	3
13	VECTOR YELLOW	AMPL.ERROR	%	0	+2	+8	+2	+10
		PHASE ERROR	DEGREES	-3	+3	0	0	+1
	CYAN	AMPL.ERROR	%	-4	-2	-5	0	+3
		PHASE ERROR	DEGREES	0	+2	+5	+3	-4
	GREEN	AMPL.ERROR	%	-2	+2	[illegible]	[illegible]	-
		PHASE ERROR	DEGREES		+2	+4	+4	0

TABLE 3-7 (Cont'd)

13 CONT'D

A	MAGENTA	AMPL.ERROR	%	+3	0	+2	0	+8
C		PHASE ERROR	DEGREES	-2	[illegible]	[illegible]	+2	-2
C								
U	RED	AMPL.ERROR	%	0	0	0	0	+2
R		PHASE ERROR	DEGREES	0	0	0	0	-4
A								
C	BLUE	AMPL.ERROR	%	-2	-3	0	-3	+5
Y		PHASE ERROR	DEGREES	-4	+2	+2	+2	+2
	MAX.	AMPL. ERROR	%	-4	-3	+8	-3	[illegible]
	MAX.	PHASE ERROR	DEGREES	-4	+3	+5	+4	-4

be very close to correct and matches the value obtained from the test tape. The standard multiburst frequencies are too much spread out and do not cover the range between 2 and 3 MHz which is important to describe codec performance.

For luminance nonlinearity and differential gain and phase, only the values at APL=50% were used since other values are either missing or questionable due to excessive noise. Two different corrected values are given for chrominance nonlinear phase. The first value is simply the result of subtraction of the figures on Table 3-2 as mentioned above. The second value (in parenthesis) was derived by first subtracting the individual phase readings and thus achieving a corrected phase error for each of the 3 subcarrier levels and then taking the maximum difference between these numbers as the final result. These values are not shown on the tables to avoid undue complexity but the second set of values is more likely to be correct. At any rate, all numbers are too small to have much impact on the final results. In vector accuracy measurements, values were computed for all 6 color bars but only the maximum error values will be used in further considerations.

Direct measurements would be the best basis for further analysis, but unfortunately they are available only for the CLI Rembrandt codec and therefore cannot be used for checking correlation between objective and subjective tests. However, by comparison with the corrected tape derived measurements, they enhance the confidence in the validity of the measured data.

Comparison with the CLI tape measurements shows reasonable agreement on many of the most important parameters though some decided differences are obvious. This was to be expected since the tests were performed on different though similar codec models. In the vector accuracy measurement comparison, the agreement on the large positive yellow amplitude error is of interest. During the 1984/85 subjective comparative evaluations, a decidedly yellow appearance of the CLI picture was frequently noticeable but apparently did not influence the scoring of the evaluators.

3.4 Motion Testing

All codec testing programs have shown clearly that for the average viewer motion rendition is the prime factor in judging codec performance. Therefore, it is highly desirable to devise an objective method to evaluate motion performance but so far this has remained an elusive goal. The DIS codec test tape contains two sequences designed especially for potential numerical motion evaluation. Both are based on the fact that a switch between two radically different pictures producing abrupt changes of many pixels simulates rapid motion. The codec output is not able to immediately follow the input change. It would be ideal if there was a measurable residue but for an initial appraisal of the concept visual observation is a sufficient practical method.

The two motion test sequences on the tape are a switch between a white "window" and a black field, and between a yellow and blue field, at 10 second intervals. Both transitions appear practically instantaneous at the codec output because the switch between such very simple images does not strain the capability of the codec algorithm. Examining the transition on the tape frame-by-frame showed no significant features of the yellow-blue switch. The white window-black field switch, however, gave a good indication that the concept is viable and with proper modifications will achieve useful results.

Following are the observations of the white window-black field transitions on the four codec output tapes.

a) GEC. The window changed over 3 frames to a mottled black which took another 77 frames to disappear completely.

b) Fujitsu. The window changed over 2 frames to a mottled black which took another 50 frames to disappear completely.

c) CLI. The first 2 frames after the switch contained fairly strong white bars which disappeared gradually after another 20 frames.

d) NEC. The window changed after one frame to a mottled black which persisted for over 6 seconds (180 frames).

Interpretation of these results is not straightforward. Though the duration of the after-image could be a measure of motion performance, the strictly visual observation cannot put a numerical value on the equally important residual amplitude. The number of frames necessary before the transition to the after-image depends also on the interpolation and frame repetition scheme of the codec and thus is not a valid measure of motion performance. However, after-image duration may be used on an interim basis until a better method is developed.

4.0 TEST DATA CORRELATION

4.1 Methodology

The available subjective and objective test data give the results in different forms and units. Correlation can be investigated only after all data have been reduced to a common denominator. It was chosen arbitrarily to normalize all test results to numbers between zero and one, with one representing the best result and zero the worst, regardless of whether in the actual measurements a higher or lower value indicates better performance. Wherever the ideal measured result is a reading of zero with possible deviations in both directions, only the absolute value of the measurement was taken into account since the direction of the deviation is generally immaterial. The mean scores of the subjective codec evaluations were treated in the same manner as the objective test results.

Table 4-1 recapitulates the pertinent data from Tables 2-2, 2-3 and 3-7 and shows the normalized values computed from them. The category of motion has been added, with the objective results based on the number of frames needed to fully complete a switched transition, as described in Paragraph 3.4. Only parameters for which complete data are available have been taken into account.

When reviewing the data it becomes apparent that not all parameters are useful in checking for data correlation. Whenever there are no or only very small objective performance differences

PARAMETER	CORRECTED VALUE GEC	FUJ	CLI	NEC	NORMALIZED VALUE GEC	FUJ	CLI	NEC
STILL & SLOW MOTION SUBJECTIVE	+.52	-1.25	+.77	+.57	.90	0	1	.32
LIVELY MOTION-SUBJECTIVE	+.65	-1.75	+1.15	-.06	.92	0	1	.59
FREQUENCY RESPONSE	240	110	200	220	1.0	0	.7	.85
CHROM/LUM GAIN INEQUALITY	12	3	7	7	0	1	.6	.6
CHROM/LUM DELAY INEQUALITY	312	165	171	0	0	.48	.45	1
LINE TIME WAVEFORM DIST.	1	1	1	1	1	1	1	1
SHORTTIME WAVEFORM DIST. (2T PULSE)	-18	-10	-19	-17	.11	1	0	.22
DIFFERENTIAL GAIN	2	1	5	1	.75	1	0	1
DIFFERENTIAL PHASE	6	2	0	3	0	.7	1	.5
CHROM/LUM INTERMOD	0	0	0	1	1	1	1	0
CHROM NON-LINEAR GAIN (TOTAL)	2	1	2	1	0	1	0	1
CHROM NON-LINEAR PHASE	2	4	3	6	1	.5	.75	0
VECTOR ACCURACY MAX AMPL. ERROR	4	3	8	3	.8	1	0	1
VECTOR ACCURACY MAX PHASE ERROR	4	3	5	4	.5	1	0	.5
MOTION (TRANSITION FRAME NO'S)	77	50	20	180	.65	.02	1.0	0

TABLE 4-1 NORMALIZED EVALUATION DATA

between the codecs, any attempt to establish correlation would yield only trivial or misleading results.

4.2 Results

The results of the correlation evaluation are shown on Figure 4-1 to 4-10. All figures have the same format. The ordinate is common to all figures and gives the normalized value of the subjective tests. Still and slow motion scenes are used throughout except for Figure 4-10 which uses the values of the lively motion tests. The abscissa gives the normalized value of the parameter for which correlation is being investigated. The point of intersection of ordinate and abscissa indicates the amount of correlation of subjective and objective evaluations for each codec. The dashed diagonal line is the locus of all points of ideal correlation.

All parameters where any amount of correlation appeared feasible were used in the evaluation. Reviewing the contents of Table 4-1, only Line Time Distortion, Chrominance/Luminance Intermodulation, and Chrominance Non-Linear Gain have been omitted, because they could not yield significant results. All other parameters have been used, producing results of varying significance and value.

4.3 Discussion

Review of the results shows much less correlation than was generally anticipated. Frequency Response (Figure 4-1) is the

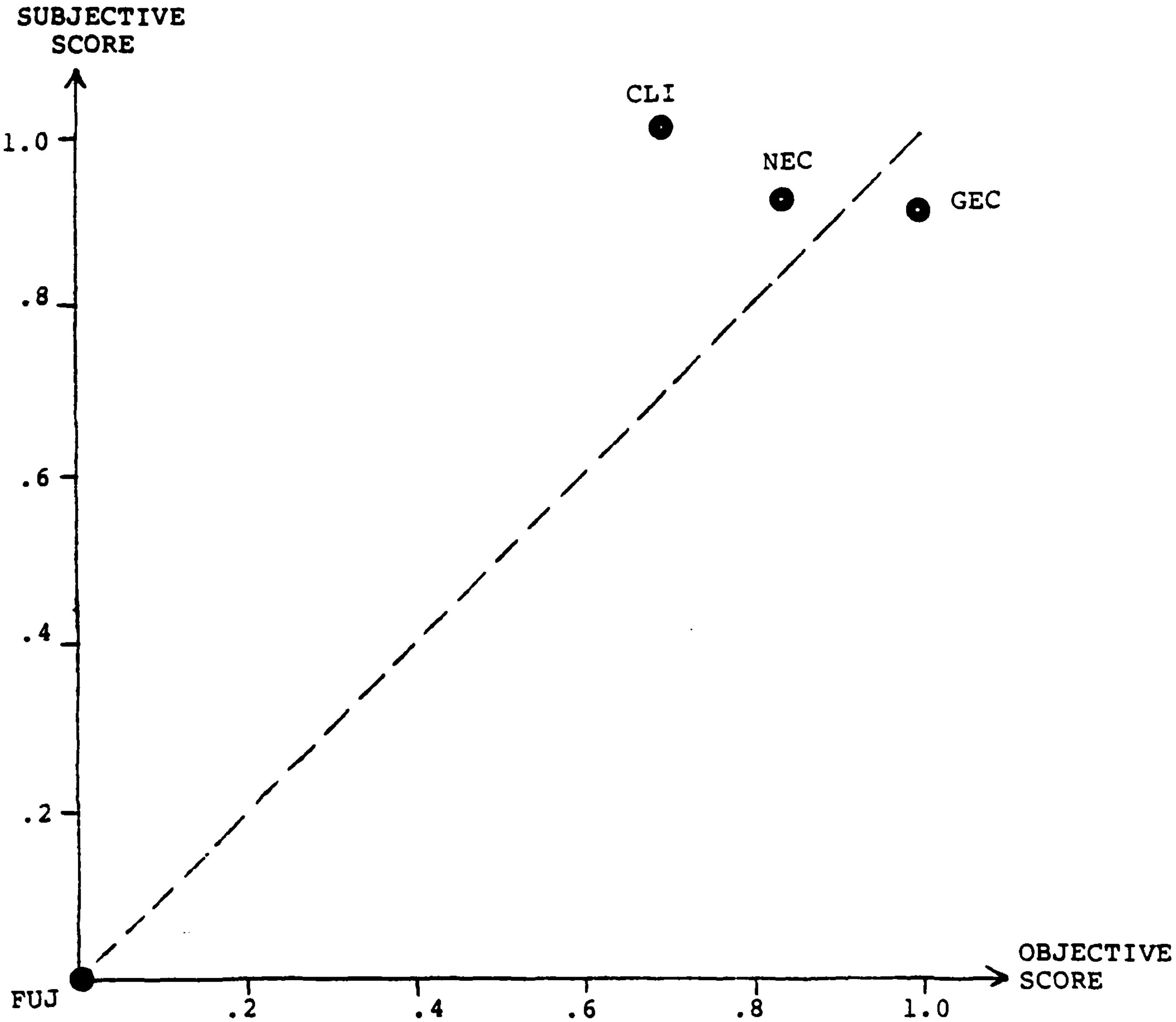


Figure 4-1

Correlation - Frequency Response

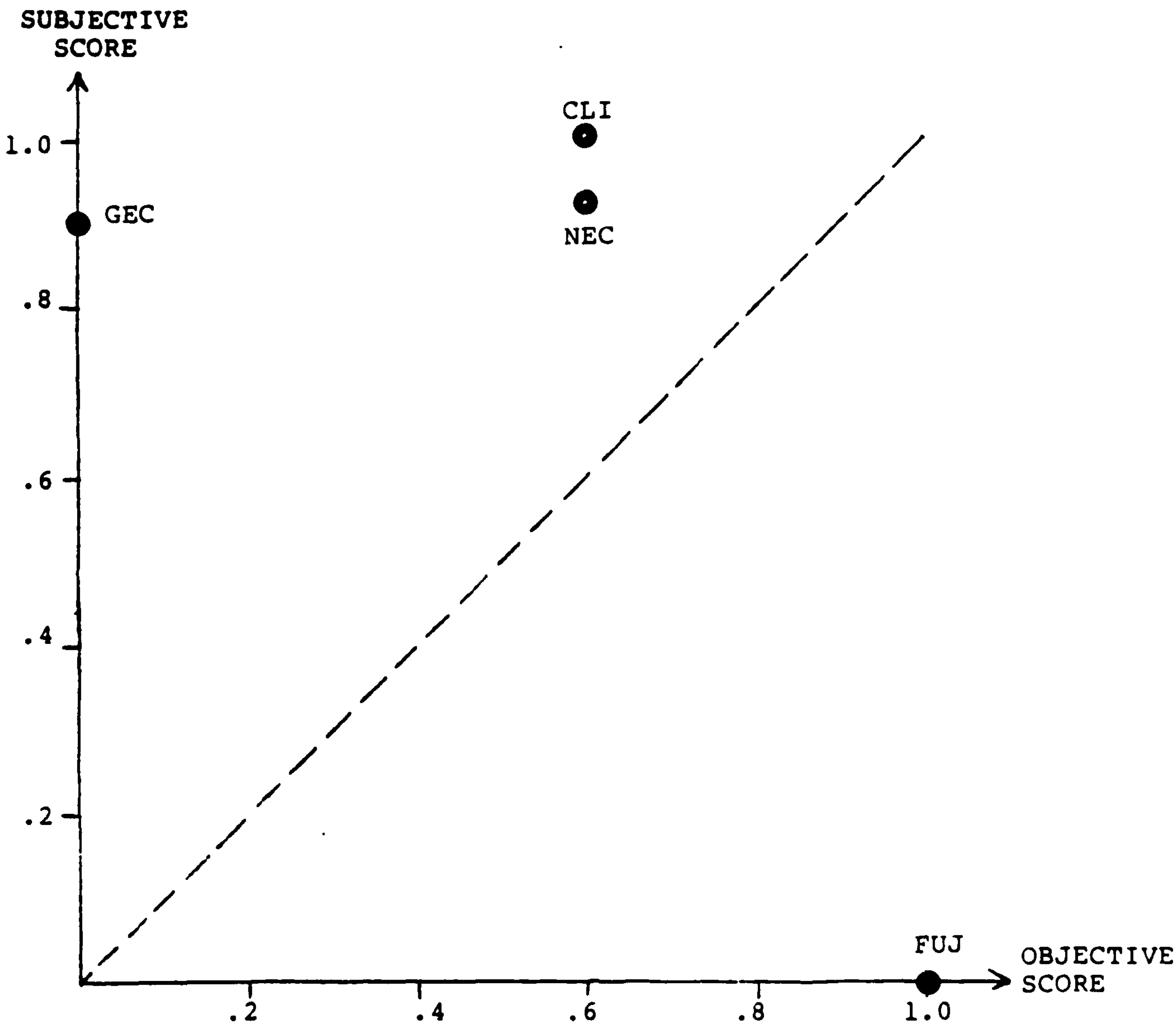


Figure 4-2

Correlation - Chrominance/Luminance Gain Inequality

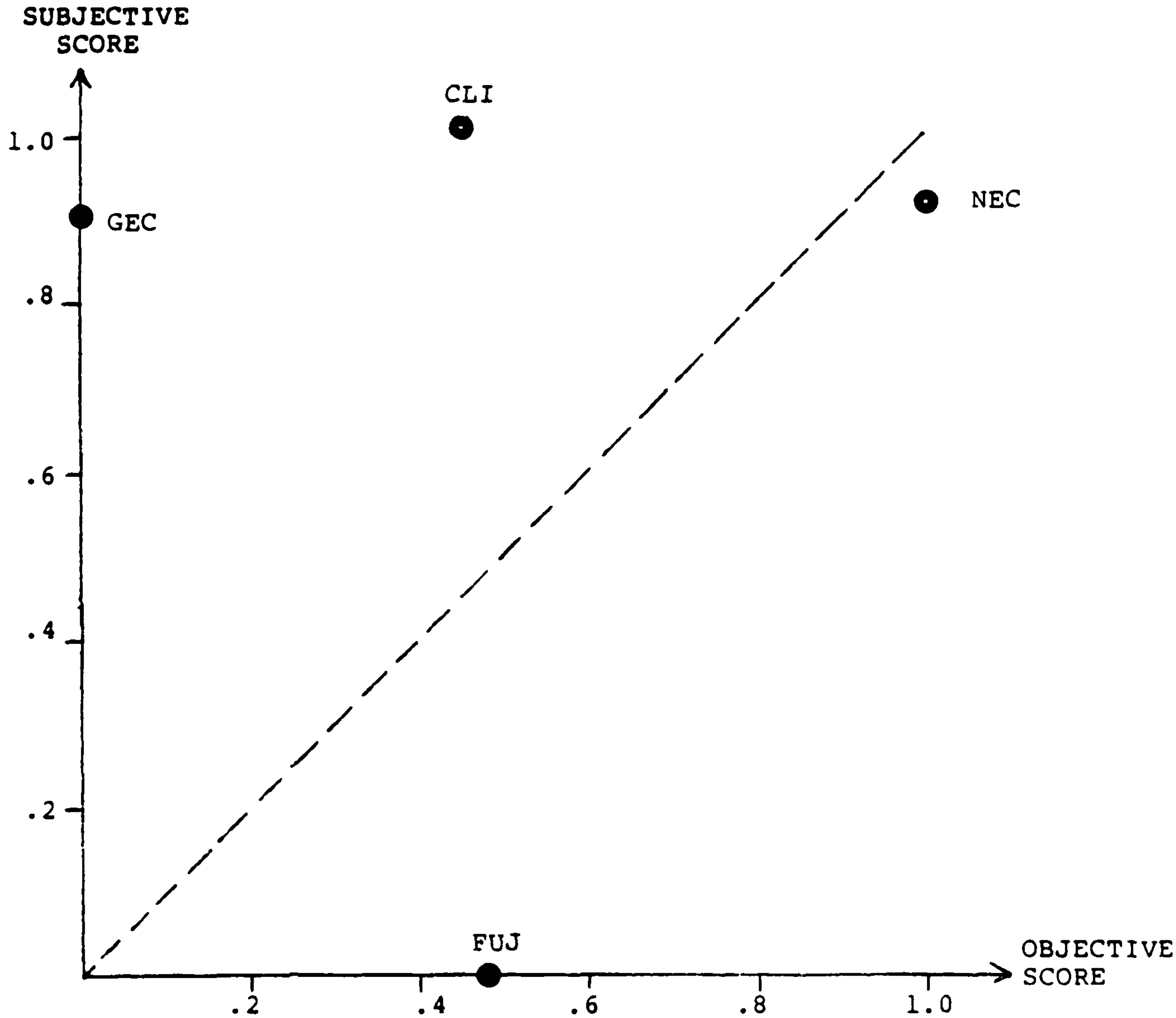


Figure 4-3

Correlation - Chrominance/Luminance
Delay Inequality

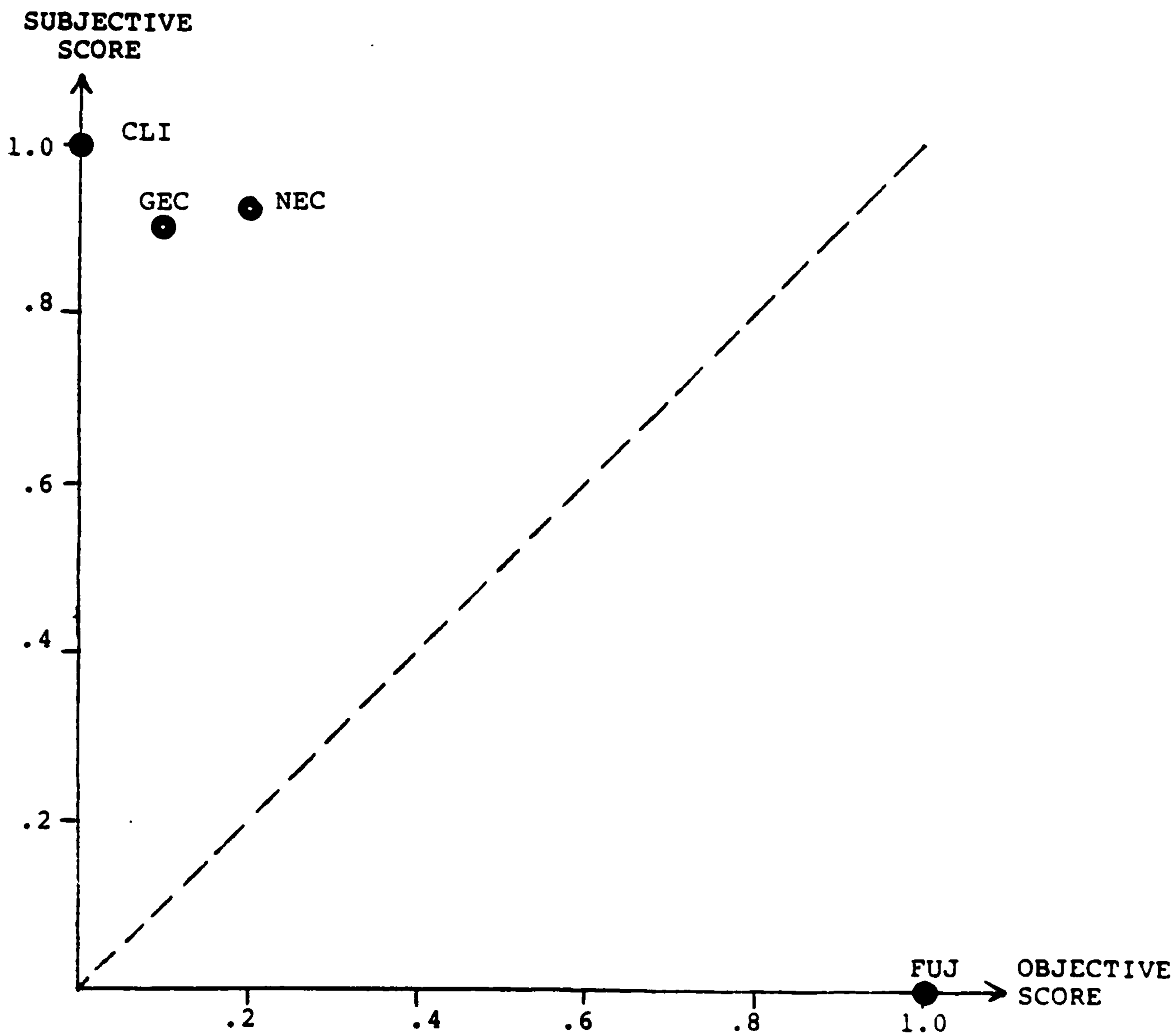


Figure 4-4

Correlation - Short Time Waveform Distortion (2T Pulse Amplitude)

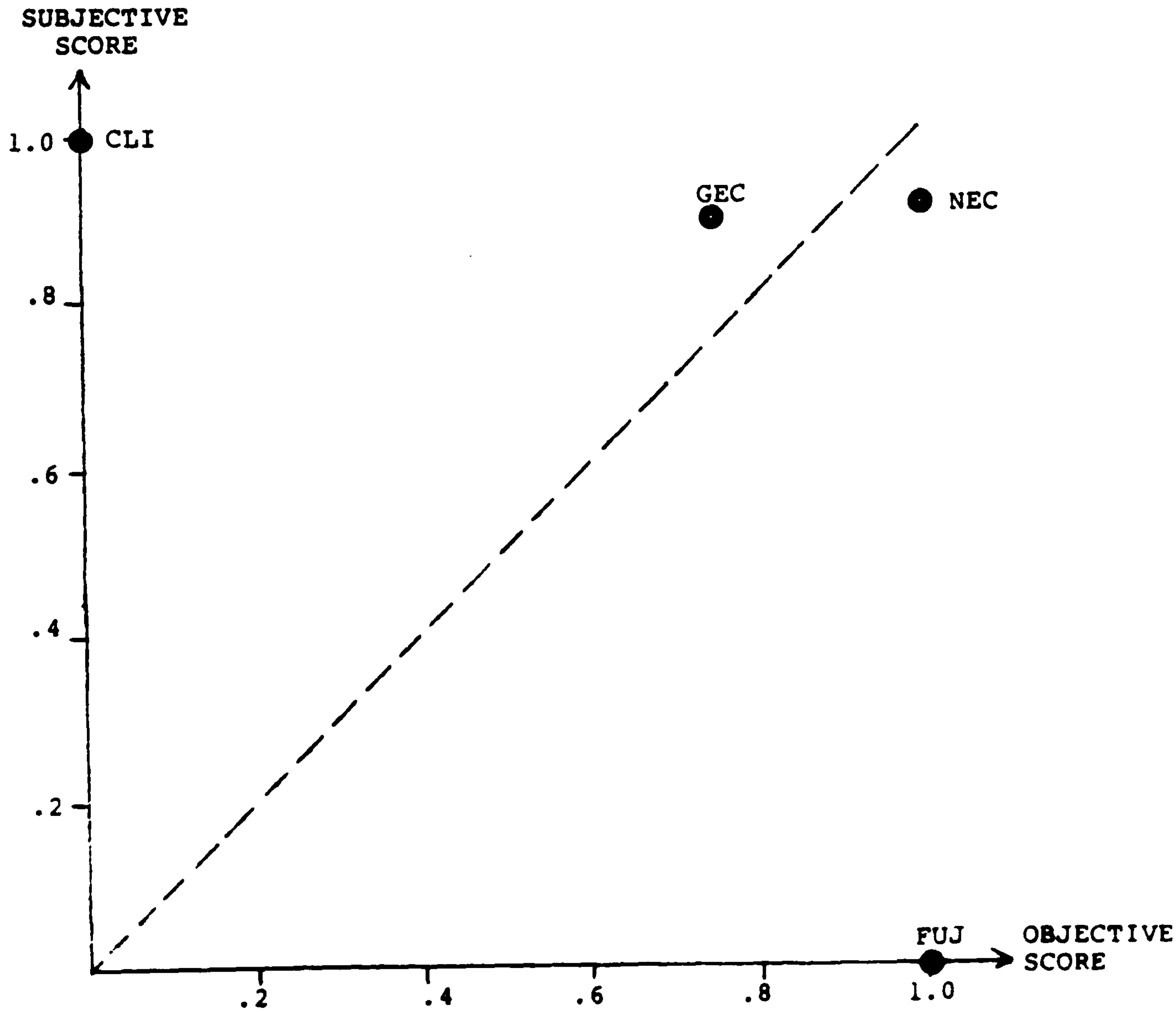


Figure 4-5

Correlation - Differential Gain

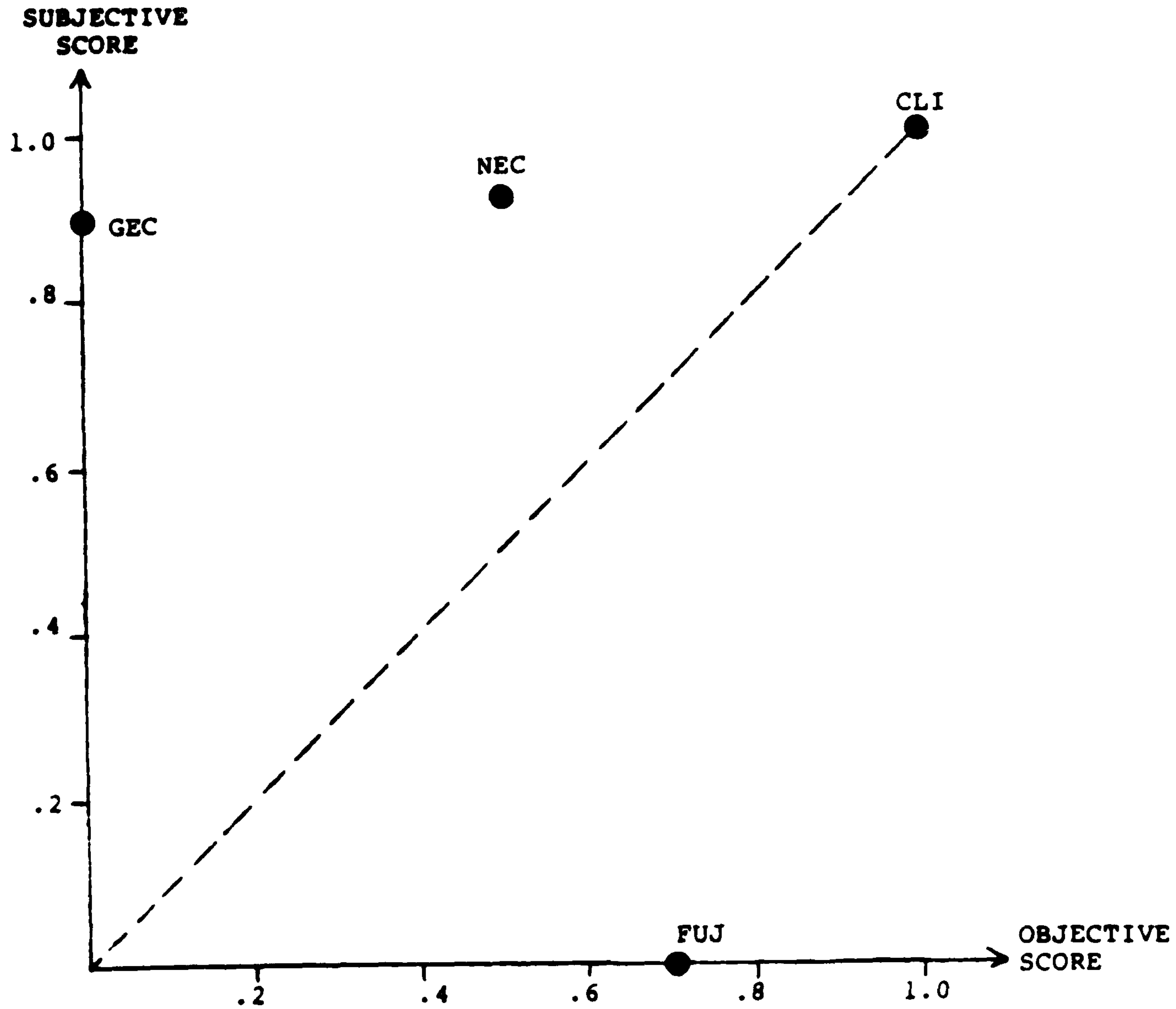


Figure 4-6

Correlation - Differential Phase

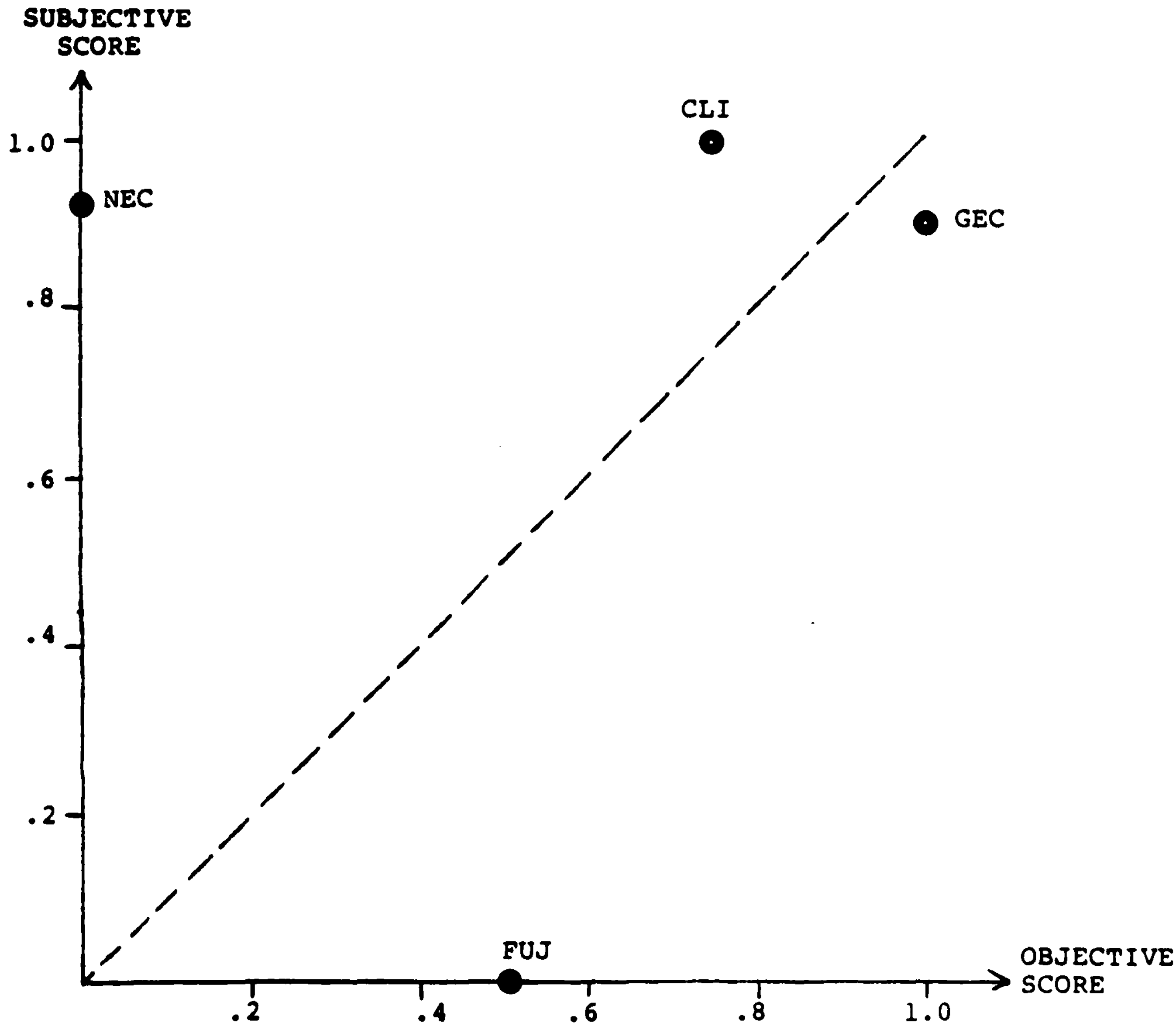


Figure 4-7

Correlation - Chominance Non-Linear Phase

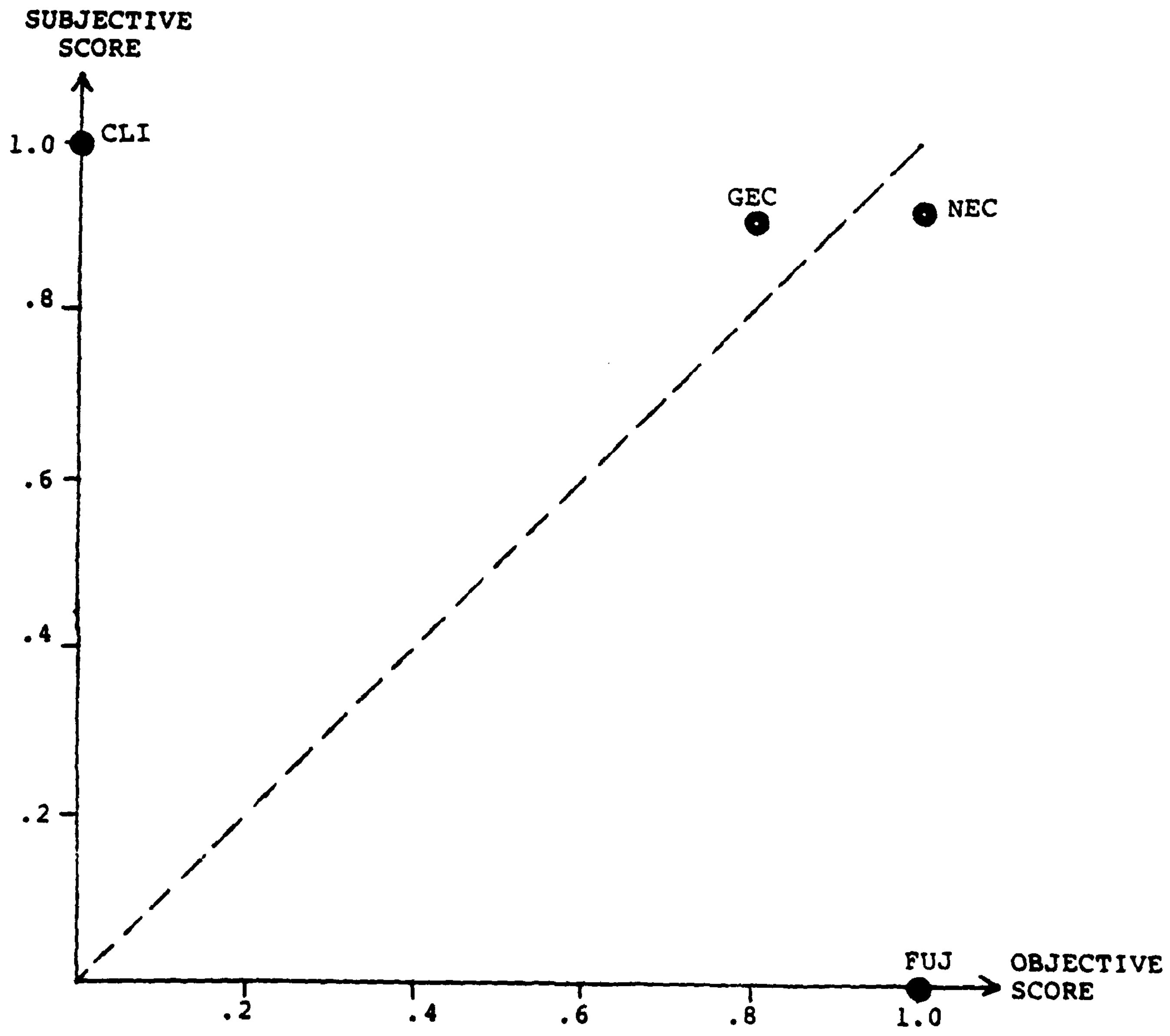


Figure 4-8

Correlation - Vector Accuracy Maximum Amplitude Error

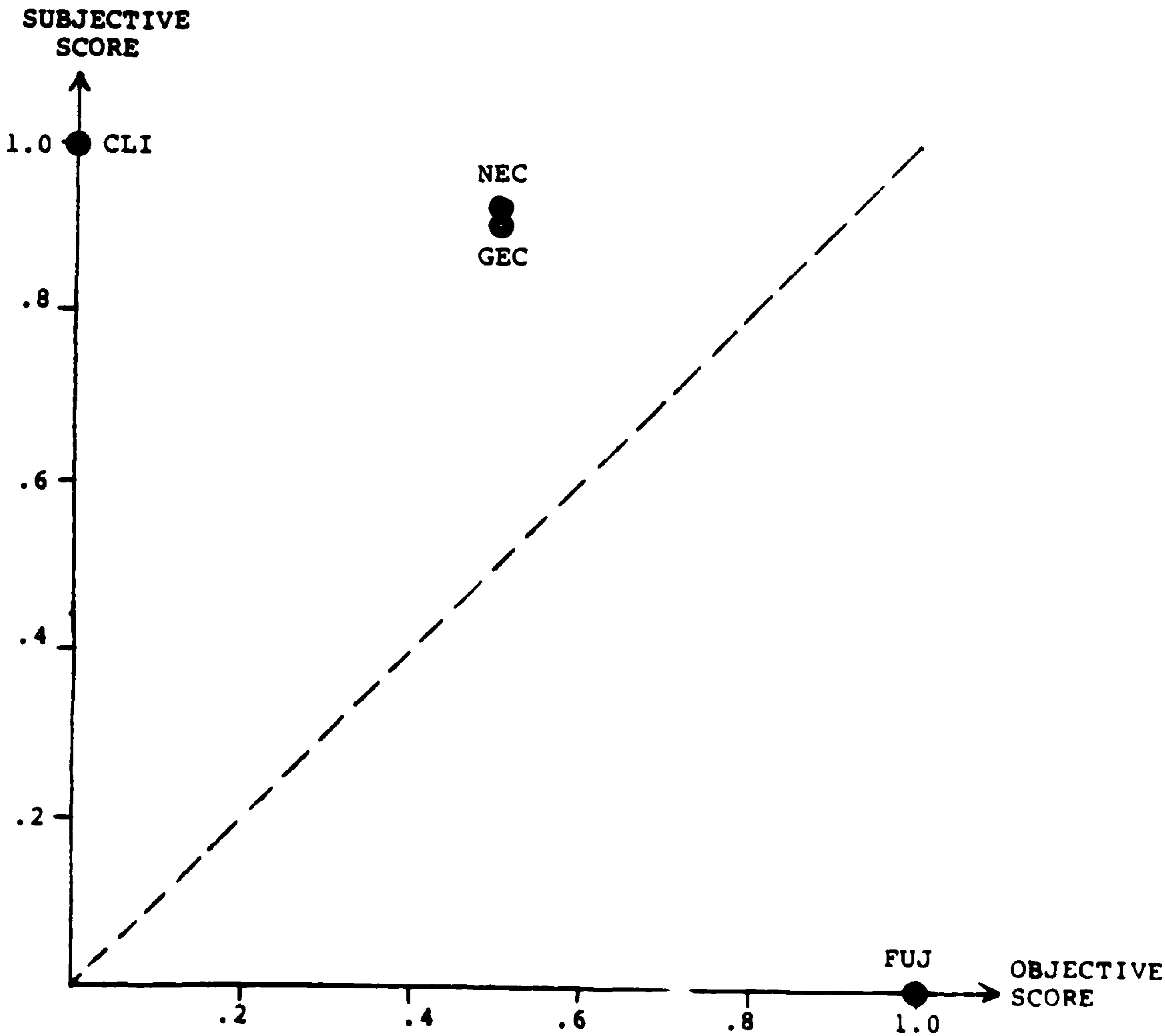


Figure 4-9

Correlation - Vector Accuracy

Maximum Phase Error

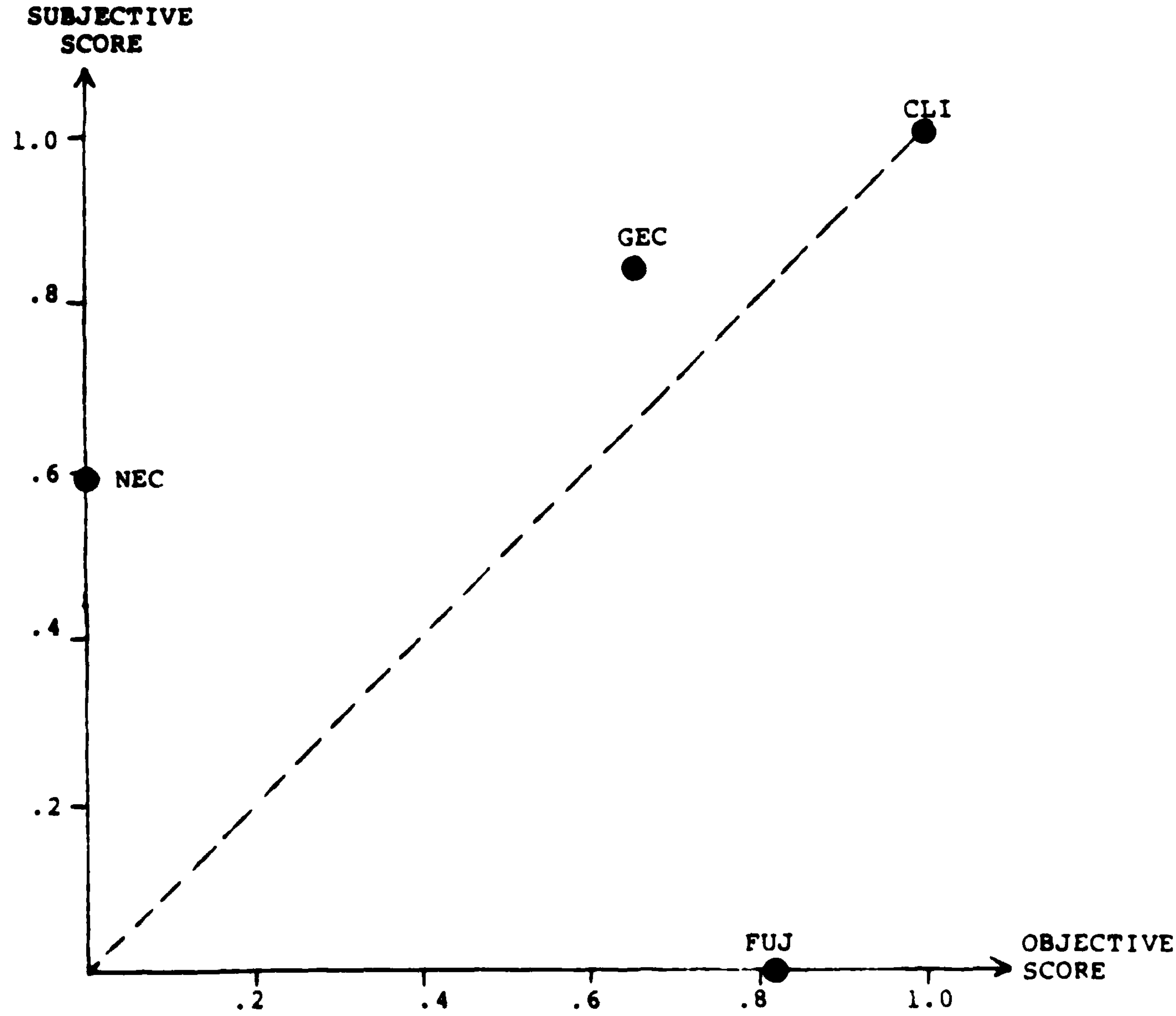


Figure 4-10

Correlation - Motion
(Number of Transition Frames)

only parameter displaying a reasonable though not ideal degree of correlation. For some of the other parameters the results for two codecs are fairly near the line of ideal correlation which however cannot be considered a significant result. In many instances the results are completely scattered.

Though more correlation was originally expected, the obtained results can be explained and are definitely useful. Frequency response is one of the most important parameters in any video transmission system, and its effect is readily visible in the reproduced picture. Furthermore, codec frequency response is always severely restricted compared to an analog system. Therefore, variations can be easily recognized even by non-expert subjective evaluators. Limitation of frequency response affects some other parameters, such as chrominance/luminance gain inequality and short time waveform distortion which makes correlation for these parameters less likely.

Most of the other parameters which are generally measured and specified in an analog video system have values within, or near to, broadcast specification limits and do not vary much between different codecs. Their values seem to be mainly determined rather randomly by incidental variations in the analog reconstitution circuitry of the decoder, and not by inherent differences in the codec algorithm. Furthermore, the performed subjective tests were not conducive to recognizing small deviations in analog performance parameters. For instance, as mentioned previously, the impairment in vector accuracy produced by the CLI codec, though visible, evidently was ignored by most evaluators. Evaluation of small differences requires comparison between original and impaired picture by video experts which is the method by which the present broadcast standards were established. This, however, was done with only slightly impaired pictures which is not the case when codec performance is to be evaluated. Unless an analog parameter is severely degraded, it is not likely to be so recognized in subjective comparative codec performance evaluation. Thus the subjective codec evaluation methodology makes good correlation between many subjective and objective measurements unlikely. It stands to reason that many parameters which are important for high quality analog or digital pictures are not significant in the evaluation of the inherently degraded outputs of digital codecs.

Five out of the ten correlation diagrams show the rather disturbing feature of complete lack of correlation, namely that the codec rated best subjectively is worst objectively, and vice versa. This makes the respective parameters poor candidates for correlation but may also largely be due to a very limited range of the objective measurement values which tends to yield trivial results. It is true that in the case of such a small range of actual measurements it may not be justifiable to normalize them over the whole range from zero to one. A range of, for instance, .3 to .7 may be more realistic and descriptive and would yield better correlation but would be completely arbitrary and could not be firmly supported.

The above does not apply to motion tests which are unique to codecs. There is no established objective test method and obviously no standard. As mentioned in Paragraph 3.4, the numerical values derived from the very limited and unsophisticated initial tests that were performed are no more than a first attempt to describe the motion capability of the codec. Therefore, even the indication of a very limited correlation shown in Figure 4-10 is an encouraging initial result. A significant improvement can be expected only after a good method for objective motion measurements has been established.

5.0 CONCLUSION AND RECOMMENDATIONS

5.1 Review of Results

The main purpose of this study was to evaluate the test signal portion of the DIS codec test tape and to establish correlation between subjective and objective results. The measurement on the processed test signals ran into some difficulties because the taped input signals had some unexplained deficiencies. Nevertheless, test results could be achieved by subtraction of measured input and output values. The validity of this method could be verified by comparison with highly reliable direct measurements on one codec only.

The correlation of subjective and objective test results was much lower than anticipated. Only the measured frequency response correlated fairly well with the subjective ranking of the codecs. Most other parameters had only rather small random variations and stayed within, or close to, standard analog performance limits. Thus these parameters were shown to be of lesser importance in describing codec performance, and had little or no correlation with subjective ranking. This is a significant result because it shows that the measured values of such parameters are not key to the objective ranking of codec performance.

Motion performance of a codec is the most important parameter in subjective evaluation. There is no established method for objective motion measurements but initial ideas were developed and incorporated in the DIS codec test tape. The results showed that the idea was feasible but that many refinements of the basic technique will be needed to establish good correlation with subjective results. Achievement of the stated goal of this program, namely elimination of the need for subjective testing, will largely depend on the availability of a good method of objective motion capability measurements.

5.2 Recommended Future Efforts

The work performed on this study presents merely a first attempt at recommending objective measurements on codecs to replace the very cumbersome subjective evaluations. It, therefore, stands to reason that not all objectives could be fulfilled but what has been accomplished clearly points the way towards necessary future efforts.

Putting test signals on tape and then processing the tape through the codec adds extra steps and, as shown by experience, potential distortions to the objective measurement process. Meanwhile, the convenience and accuracy of direct measurements with modern equipment has been demonstrated. It, therefore, is recommended to make measurements on all the latest design codecs (possibly at more than one data rate) using a test setup similar to the one on Figure 3-7. This process will provide more than the necessary parameter measurements without extra effort. However, care will have to be taken that the test signals are adapted to the limitations of the codec. One signal that definitely requires modification is the multiburst which must be limited to the range up to a maximum of 3 MHz and contain frequency packets in the (for a codec) very critical range between 2 and 3 MHz. It is understood that such a modification of the Tektronix 1910 Signal Generator can be accomplished readily by reprogramming and replacing one PROM.

Other test signals not commonly used for analog video have found acceptance in the evaluation of high data rate digital PCM and DPCM systems. They are a steep rise horizontal rate step function, a ramp with a low variable slope and variable setup, and a flat field with variable setup. These signals are used to determine various quantizing distortions. The additional extensive processing in low data rate digital codecs may often completely overshadow these distortions but tests with such signals are highly recommended. The signals are either directly available in the Tektronix 1910 Generator or can be produced with minimal modifications.

It has been shown by many subjective tests that motion rendition is the most critical codec performance parameter. Up to now no method of objectively measuring motion performance has been established. This program presents a first effort in this direction. The accomplished results have shown distinct differences between codecs but more varied and complex patterns will be needed to produce accurate and consistent numerical values. It is recommended that a program be initiated to establish patterns (in the simplest form possibly various size checker boards) which can demonstrate the difference between codecs not only visually but also allow integration and measurement of the residual signal after switching and thus may provide a numerical value describing codec motion performance. The test pattern(s) will have to be chosen with great care such as to not favor any particular algorithm. A successful program in this area will make it possible to reduce the lengthy subjective evaluation to a brief objective measurement.

END

9-87

DTIC